

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2003, 08:53:46 ; Search time 67 Seconds

(Without alignments)
1020.262 Million cell updates/sec

Title: US-09-502-426a-2

Perfect score: 2681

Sequence: 1 MEETENHTLLPLLLPSLTS.....FAFPYDFNGPLFRVSRIL 513

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A_Geneseq_101002:.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2681	100.0	513	21	AA07921
2	978.5	36.5	472	18	AA077153
3	978.5	36.5	472	21	AA044571
4	978.5	36.5	472	21	AA045022
5	978.5	36.5	491	21	AA045021
6	978.5	36.5	492	21	AA044570
7	946	35.3	444	21	AA044572
8	946	35.3	444	21	AA045023
9	866	32.3	512	23	AB01571
10	816.5	30.5	486	22	AA002839

11	726	27.1	461	21	AA046490
12	726	27.1	462	21	AA046489
13	726	27.1	465	21	AA046491
14	712	26.6	461	21	AA011836
15	712	26.6	465	21	AA011835
16	712	26.6	465	21	AA011834
17	653.5	24.4	468	21	AA020763
18	606	22.6	471	21	AA033013
19	606	22.6	479	21	AA033012
20	603	22.5	469	21	AA033014
21	598.5	22.3	430	21	AA020764
22	597.5	22.3	485	22	AA002835
23	596.5	22.2	497	22	AA002829
24	592	22.1	496	22	AA002837
25	588	21.9	509	22	AA002832
26	585	21.8	507	22	AA002830
27	583	21.7	484	22	AA002822
28	571	21.3	501	22	AA002823
29	569.5	21.2	493	22	AA002836
30	568.5	21.2	500	22	AA002824
31	564	21.0	484	22	AA002821
32	564	21.0	500	22	AA002826
33	558.5	20.8	483	22	AA002827
34	558.5	20.8	512	22	AA002834
35	553.5	20.6	498	22	AA002838
36	547.5	20.4	489	21	AA030048
37	546	20.4	481	21	AA030049
38	534.5	19.9	503	22	AA002828
39	528.5	19.7	388	21	AA020785
40	503	18.8	492	19	AA037733
41	503	18.8	492	19	AA044159
42	503	18.8	492	22	AA085155
43	503	18.3	492	23	AA015325
44	490	18.3	433	21	AA030050
45	461	17.2	497	19	AA037735

ALIGNMENTS

RESULT 1	
AA07921	standard; Protein: 513 AA.
AA07921:	
14-NOV-2000	(first entry)
XX	
AC	AA07921:
XX	
DT	14-NOV-2000 (first entry)
XX	
DE	A cytochrome P450 enzyme designated DMF4.
XX	
KW	DMF4: cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
KM	plant phenotype; cell elongation.
XX	
OS	Arabidopsis sp.
PN	W0200047715-A2.
XX	
PD	17-AUG-2000.
XX	
PF	11-FEB-2000; 2000WO-US03820.
XX	
PR	11-FEB-1999; 99US-0119657.
XX	
PR	11-FEB-1999; 99US-0119658.
XX	
FA	(ARIZ-) ARIZONA BOARD OF REAGENTS.
XX	
PI	Azpiroz R, Choe S, Feldmann KA;
XX	
DR	WPI: 2000-549142/50.
XX	
DR	N-PSDB: AAA59599.
XX	
PT	New isolated dwf4 polynucleotide useful for altering the phenotype of plants, for diagnostic assays and in the production of antibodies -

XX Claim 50; Fig 11; 113pp; English.
 PS
 CC The present sequence represents a DMF4 polypeptide. The polypeptide is a
 CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
 CC brassinosteroids. Specifically, it mediates multiple
 CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DMF4
 CC polynucleotide is used for altering the phenotype of a plant. DMF4
 CC plants display a dramatic reduction in the length of different organs,
 CC and this reduction is attributable to a defect in cell elongation.
 CC The DMF4 polynucleotides and polypeptides can be used in diagnostic
 CC assays and to generate antibodies, which can be used to produce
 CC immunogenic compositions.
 CC
 XX
 SQ Sequence 513 AA:

Query Match 100.0%; Score 2681; DB 21; Length 513;
 Best Local Similarity 100.0%; Pred. No. 1.3e-228;
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 METEHTHTLLPLLLPSLSLSLFLILKRRNRKTRFNPFGKSGMPFAGETIGYAKPYT 60
 DB 1 METEHTHTLLPLLLPSLSLSLFLILKRRNRKTRFNPFGKSGMPFAGETIGYAKPYT 60
 QY 61 ATTLDGFMQOHVSKYKIRSNLFGEPITVSADAGLNRETLQNEGLFECSTPRSTIGIL 120
 DB 61 ATTLDGFMQOHVSKYKIRSNLFGEPITVSADAGLNRETLQNEGLFECSTPRSTIGIL 120
 QY 121 GKMSMLVGVGDHNRDMRSISLNFSLSHARLRTILKDVERTFLVDSMOQNSIFSADNER 180
 DB 121 GKMSMLVGVGDHNRDMRSISLNFSLSHARLRTILKDVERTFLVDSMOQNSIFSADNER 180
 QY 181 KFTFTPLAKHIMSMDPGEETBOLKEVYTFMKGVVAPLNPCTAYHKALOSRTIILK 240
 DB 181 KFTFTPLAKHIMSMDPGEETBOLKEVYTFMKGVVAPLNPCTAYHKALOSRTIILK 240
 QY 241 FTERKKEERKLDIKERDQEEVEKTEDEAKMSKSDHVRKORDDDLLGWLKHSNSTEQ 300
 DB 241 FTERKKEERKLDIKERDQEEVEKTEDEAKMSKSDHVRKORDDDLLGWLKHSNSTEQ 300
 QY 301 IIDLILSLFAGHETSSVALAIFFLQACPAVEELREHLEIARAKKELGESELMWMD 360
 DB 301 IIDLILSLFAGHETSSVALAIFFLQACPAVEELREHLEIARAKKELGESELMWMD 360
 QY 361 YKKMOTQOVINETLRLGVNVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLDSRYD 420
 DB 361 YKKMOTQOVINETLRLGVNVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLDSRYD 420
 QY 421 QPNLFPNPMQOONNGASSSGSSEFTWGNNTMPFGGGRCLCAGSELAKLEMAVFIHILV 480
 DB 421 QPNLFPNPMQOONNGASSSGSSEFTWGNNTMPFGGGRCLCAGSELAKLEMAVFIHILV 480
 QY 481 LKFNWELADDDQPFAPFPVDFPNGLPIRVSRIL 513
 DB 481 LKFNWELADDDQPFAPFPVDFPNGLPIRVSRIL 513

RESULT 2
 AAM27153
 ID AAM27153 standard; Protein; 472 AA.

XX AAM27153;
 AC
 XX
 DT 14-APR-1998 (first entry)
 XX
 DE Arabidopsis thaliana cytochrome P450-type hydroxylase.
 XX
 KW Cytochrome P450-type hydroxylase; identification; brassinosteroid;
 XX brassinosteroid inhibitor; modified plant; recombinant production;
 XX teasterone.
 OS Arabidopsis thaliana.
 XX

PN W09735986-A1.
 XX
 PD 02-OCT-1997.
 XX
 PF 27-MAR-1997; 97WO-EP01586.
 XX
 PR 27-MAR-1996; 96US-0622166.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Altman T, Koncz C, Mathur J, Szekeres MA;
 XX
 DR WPI: 1997-489649/45.
 DR N-PSDB: AAT85306, AAT85307.
 XX
 PT New isolated plant cytochrome P450-type hydroxylase gene - used to
 PT identify substances acting as brassino-steroid(s) or brassinosteroid
 PT inhibitors for the production of modified plants
 PS
 PS Claim 1; Pages 44-46; 77pp; English.

CC The present sequence is Arabidopsis thaliana cytochrome
 CC P450-type hydroxylase. The hydroxylase can be used to identify
 CC brassinosteroids or brassinosteroid inhibitors, useful to produce
 CC plants with modified physiological and/or phenotypic
 CC characteristics. The modified plants may show, e.g. stimulation of
 CC growth, increased cell elongation, increased wood production,
 CC accelerated seed germination at low temperatures, an increase in
 CC dry weight, repressed anthocyanin production during growth in light
 CC and/or inhibited de-etiolation which is induced, e.g. by cytokinin,
 CC in the dark or an increase in stress tolerance. The hydroxylase or
 CC its coding sequence can also be used for the recombinant production
 CC of compounds, e.g. teasterone.
 CC
 XX

SQ Sequence 472 AA:

Query Match 36.5%; Score 978.5; DB 18; Length 472;
 Best Local Similarity 41.9%; Pred. No. 6.6e-78;
 Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSLSLFLILKRRNRKTRFNPFGKSGMPFAGETIGYAKPYTATTLGDFMOOH 71
 DB 7 LLLSSIAAGPL---LLLRTRRRKGLPGLSGPLIETQLGAYTEHNEPPIDDR 63
 QY 72 VSKYKITYSNLFGEPITVSADAGLNRETLQNEGLFECSTPRSTIGILGKMSMLVGVGD 131
 DB 64 VARGSGVFHTLFGEPITVSADAGLNRETLQNEGLFECSTPRSTIGILGKMSMLVGVGD 123
 QY 132 MHRDMRSISLNFSLSHARLRTILKDVERTFLVDSMOQNSIFSADNERKFTFPLMAKH 191
 DB 124 LKRNHSLTMSFRANSSTIYDHLMDLIDRLVRRNLDSSSRVLL--MEKAKTTFELTVAK 181
 QY 192 INSMQGEETBOLKEVYTFMKGVVAPLNPCTAYHKALOSRTIILKFERKMEERL 251
 DB 182 LMSFDPG-EMSSSLRKREYLVIEGFPPLPLPSTYTRAIQAR-----RKVAEALT 232
 QY 252 DIKEDEQEEVEKTEDEAKMSKSDHVRKORDDDLLGWLKHSNSTEQIIDLILSLFLA 311
 DB 233 VVYMKRREDEEGAE-----RKDDMLAALLAADDGSEDEIVDFVALVA 278
 QY 312 GHETSSVALAIFFLQACPAVEELREHLEIARAKKELGESELMWMDYKKMDFPQCVI 371
 DB 279 GYETSTITMTLAVKFLTEPFLALAOULKEHEKIRANKSD--SYLSEMSYKSNPFTQCVV 336
 QY 372 NETLRLGVNVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLDSRYDQPNLFPNPMQ 431
 DB 337 NETLRLGVNVRFLHRAKALKDVRKYGDIPSGMKVLPVISAHLDSRYDQPNLFPNPMQ 396
 QY 432 QONNGASSSGSSEFTWGNNTMPFGGGRCLCAGSELAKLEMAVFIHILVLFNWLAD 490
 DB 397 -----SNSVTTGSPNVEFTFGGGRCLCAGSELAKLEMAVFIHILVLTGFSWVPAED 446
 QY 491 DQPFAPFPVDFPNGLPIRVSR 511

DB 447 DKLVFFPTTRQKRYPIFYKR 467

RESULT 3
AAG44571
ID AAG44571 standard; Protein; 472 AA.
XX AAG44571;
AC
XX
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 55847.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI03405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

```

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156599.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161930.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 36.5% Score 978.5; DB 21; Length 472;
 Best Local Similarity 41.9% Pred. No. 6.6e-76;
 Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

```

QY 12 LLLPSTLSLFLTLRRNRKFRFNLPKSGWPFGEFTGYLKPATATTLGDFMOOH 71
DB 7 LLLLSSTIAGFL--LLRRRTYRRMGSLPPSSLSGLPLGTGTFQOLIGAKTKTENPEPFIDR 63
QY 72 VSKYKGIYRSMIFGEPTIVSADAGINRFTLQNEGRLFCFCSYPRSTIGTILGRKMSMLVYGD 131
DB 64 VARYGSVMTLFGPFTFSADPENTNRFVLQNEGKLFCSYPASICNLLGKHSLLMKGS 123

```

```

QY 133 MHRDRSISLNFSLHARLFTLLKQVERHTLFLVDSMOONSIFSAODEAKKFTFNLMAKH 191
DB 124 LHKRHSISLTMSPFANSIIIDHMLDIDRLVRFNLSWSSRVL--MEAKKITFELTVQ 181
QY 192 IKSMDPGEETEQLKKEYTFEMKGVYSAFLNPGTAHYHKALOSRAITLKFIRKKMEERL 251
DB 182 LMSFDPG-EWSLSLKRKLYLVTEGFFSLPLPSTTYRAIQR-----RKVAELT 232
QY 252 DIKEEDOEVEEVKTEDEAKMSKSDVHRKQRTDDDLGWLKHSNLSSTEQIIDLILSLFA 311
DB 223 VVVMKRREEEGEAE-----RKDKMLAALLAADGSDDEIYDFVALVA 278
QY 312 GHETSSVAIALAIFFLQACPRAVEELREHFLIAKAKELGSELMNDYKMDTQCVY 371
DB 279 GYETSTIMTLAVKFLTEPLPALQLEHEKERIRAKSD--SYSEWSYKSNPFTQCVY 336
QY 372 NETLRGNVVRFLHKKALKDVRKGDIPSGKVLPVYSAVLDNSRDOPLFNPWRQ 431
DB 337 NETLRVANIIGVFRAMTDVEIKGKIPKMKVSSFRRAVLDNHFEDARTFNPWRQ 396
QY 432 QONNGASSSGSFSSTWGN--YMPFGGPRLCAGSELAKLENAVFTHLVLKFMNELAD 490
DB 397 -----SNSVTGSPSNVFTPFGGGPRLCGYELARVALSVLHLRYTGFSVPRADQ 446
QY 491 DQFAFPVDFPNCGLPIKYSR 511
DB 447 DLVFEPTTRQKRYPFVKR 467

```

RESULT 4

AA645022 standard; Protein; 472 AA.

AA645022;

18-OCT-2000 (first entry)

Arabidopsis thaliana protein fragment SEQ ID NO: 56469.

Protein identification; signal transduction pathway; metabolic pathway;
 hybridisation assay; genetic mapping; gene expression control; promoter;
 termination sequence.

Arabidopsis thaliana.

Ep1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

```

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

```


PR 11-MAY-1999;	9905-0134256.
PR 14-MAY-1999;	9905-0134218.
PR 14-MAY-1999;	9905-0134219.
PR 14-MAY-1999;	9905-0134221.
PR 14-MAY-1999;	9905-0134320.
PR 18-MAY-1999;	9905-0134370.
PR 19-MAY-1999;	9905-0134941.
PR 20-MAY-1999;	9905-0135124.
PR 21-MAY-1999;	9905-0135353.
PR 24-MAY-1999;	9905-0135629.
PR 25-MAY-1999;	9905-0136021.
PR 27-MAY-1999;	9905-0136392.
PR 28-MAY-1999;	9905-0136782.
PR 01-JUN-1999;	9905-0137222.
PR 03-JUN-1999;	9905-0137528.
PR 04-JUN-1999;	9905-0137502.
PR 07-JUN-1999;	9905-0137724.
PR 08-JUN-1999;	9905-0138094.
PR 10-JUN-1999;	9905-0138540.
PR 10-JUN-1999;	9905-0138847.
PR 14-JUN-1999;	9905-0139119.
PR 16-JUN-1999;	9905-0139452.
PR 16-JUN-1999;	9905-0139453.
PR 17-JUN-1999;	9905-0139492.
PR 18-JUN-1999;	9905-0139454.
PR 18-JUN-1999;	9905-0139455.
PR 18-JUN-1999;	9905-0139456.
PR 18-JUN-1999;	9905-0139457.
PR 18-JUN-1999;	9905-0139458.
PR 18-JUN-1999;	9905-0139459.
PR 18-JUN-1999;	9905-0139460.
PR 18-JUN-1999;	9905-0139461.
PR 18-JUN-1999;	9905-0139462.
PR 18-JUN-1999;	9905-0139463.
PR 18-JUN-1999;	9905-0139750.
PR 18-JUN-1999;	9905-0139763.
PR 21-JUN-1999;	9905-0139817.
PR 22-JUN-1999;	9905-0139899.
PR 23-JUN-1999;	9905-0140353.
PR 23-JUN-1999;	9905-0140354.
PR 24-JUN-1999;	9905-0140695.
PR 28-JUN-1999;	9905-0140823.
PR 29-JUN-1999;	9905-0140991.
PR 30-JUN-1999;	9905-0141287.
PR 01-JUL-1999;	9905-0141442.
PR 01-JUL-1999;	9905-0141254.
PR 02-JUL-1999;	9905-0142055.
PR 06-JUL-1999;	9905-0142390.
PR 08-JUL-1999;	9905-0142803.
PR 09-JUL-1999;	9905-0142920.
PR 12-JUL-1999;	9905-0142977.
PR 13-JUL-1999;	9905-0143542.
PR 14-JUL-1999;	9905-0143624.
PR 15-JUL-1999;	9905-0144005.
PR 16-JUL-1999;	9905-0144085.
PR 16-JUL-1999;	9905-0144086.
PR 19-JUL-1999;	9905-0144325.
PR 19-JUL-1999;	9905-0144331.
PR 19-JUL-1999;	9905-0144332.
PR 19-JUL-1999;	9905-0144333.
PR 19-JUL-1999;	9905-0144334.
PR 19-JUL-1999;	9905-0144335.
PR 20-JUL-1999;	9905-0144352.
PR 20-JUL-1999;	9905-0144632.
PR 20-JUL-1999;	9905-0144684.
PR 21-JUL-1999;	9905-0144814.
PR 21-JUL-1999;	9905-0145086.
PR 21-JUL-1999;	9905-0145088.
PR 22-JUL-1999;	9905-0145085.
PR 22-JUL-1999;	9905-0145087.
PR 22-JUL-1999;	9905-0145089.
PR 22-JUL-1999;	9905-0145192.
PR 23-JUL-1999;	9905-0145145.
<hr/>	
PR 23-JUL-1999;	9905-0145218.
PR 23-JUL-1999;	9905-0145224.
PR 26-JUL-1999;	9905-0145276.
PR 27-JUL-1999;	9905-0145913.
PR 27-JUL-1999;	9905-0145918.
PR 27-JUL-1999;	9905-0145919.
PR 28-JUL-1999;	9905-0145951.
PR 02-AUG-1999;	9905-0146386.
PR 02-AUG-1999;	9905-0146388.
PR 02-AUG-1999;	9905-0146389.
PR 03-AUG-1999;	9905-0147038.
PR 04-AUG-1999;	9905-0147204.
PR 04-AUG-1999;	9905-0147302.
PR 05-AUG-1999;	9905-0147192.
PR 05-AUG-1999;	9905-0147260.
PR 06-AUG-1999;	9905-0147303.
PR 06-AUG-1999;	9905-0147416.
PR 09-AUG-1999;	9905-0147493.
PR 09-AUG-1999;	9905-0147935.
PR 10-AUG-1999;	9905-0148171.
PR 11-AUG-1999;	9905-0148319.
PR 12-AUG-1999;	9905-0148341.
PR 13-AUG-1999;	9905-0148565.
PR 13-AUG-1999;	9905-0148684.
PR 16-AUG-1999;	9905-0149368.
PR 17-AUG-1999;	9905-0149175.
PR 18-AUG-1999;	9905-0149426.
PR 20-AUG-1999;	9905-0149722.
PR 20-AUG-1999;	9905-0149723.
PR 20-AUG-1999;	9905-0149929.
PR 23-AUG-1999;	9905-0149902.
PR 23-AUG-1999;	9905-0149930.
PR 25-AUG-1999;	9905-0150566.
PR 26-AUG-1999;	9905-0150884.
PR 27-AUG-1999;	9905-0151065.
PR 27-AUG-1999;	9905-0151066.
PR 27-AUG-1999;	9905-0151080.
PR 30-AUG-1999;	9905-0151303.
PR 31-AUG-1999;	9905-0151438.
PR 01-SEP-1999;	9905-0151930.
PR 07-SEP-1999;	9905-0152363.
PR 10-SEP-1999;	9905-0153070.
PR 13-SEP-1999;	9905-0153758.
PR 15-SEP-1999;	9905-0154018.
PR 16-SEP-1999;	9905-0154039.
PR 20-SEP-1999;	9905-0154779.
PR 22-SEP-1999;	9905-0155139.
PR 23-SEP-1999;	9905-0155486.
PR 24-SEP-1999;	9905-0155659.
PR 28-SEP-1999;	9905-0156458.
PR 29-SEP-1999;	9905-0156596.
PR 04-OCT-1999;	9905-0157117.
PR 05-OCT-1999;	9905-0157753.
PR 06-OCT-1999;	9905-0157865.
PR 07-OCT-1999;	9905-0158029.
PR 08-OCT-1999;	9905-0158232.
PR 12-OCT-1999;	9905-0158369.
PR 13-OCT-1999;	9905-0159293.
PR 13-OCT-1999;	9905-0159294.
PR 13-OCT-1999;	9905-0159295.
PR 14-OCT-1999;	9905-0159329.
PR 14-OCT-1999;	9905-0159330.
PR 14-OCT-1999;	9905-0159331.
PR 14-OCT-1999;	9905-0159637.
PR 14-OCT-1999;	9905-0159638.
PR 18-OCT-1999;	9905-0159584.
PR 21-OCT-1999;	9905-0160741.
PR 21-OCT-1999;	9905-0160767.
PR 21-OCT-1999;	9905-0160768.
PR 21-OCT-1999;	9905-0160770.
PR 21-OCT-1999;	9905-0160814.
PR 21-OCT-1999;	9905-0160815.
PR 22-OCT-1999;	9905-0160980.

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161923.
PR 29-OCT-1999; 99US-0162142.

Query Match 36.5%; Score 978.5; DB 21; Length 472;
Best Local Similarity 41.9%; Pred. No. 6,6e-78;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLLSLLSLFLLLKRRNKRFTNLPKSGSMPFGETIGYKPYATILSGEMOQH 71
DB 7 LLLLSIANGFL--LLKRTYRRNGLPKSLGLPGLGETYQLIGAVKTENPEPPIDR 63
QY 72 VSKYKIVRSNLFGEPTIYSADAGINRFLONEGRLEFECSPRSIGILGKMSMLVYGD 131
DB 64 VARYGSVFETHLFGEPTIFSDPETNRFLVONEKLEFECSPASICNLGKHSLLMGKS 123
QY 132 MHRDRSISLNFSLHARLFTILKDYERHTLVLDWQONSIFSAODEAKKFTENIMAKH 191
DB 124 LHKRHSLSLMSFANSIIRKDHMLDIDRLVRFNLDMSRSRVLL--MEBAKKITFELTVQ 181
QY 192 IMSMDGEEETQOLKKEYTFMKGVVSAPLNPGTAHYKALOSRATILKFIKKMEERL 251
DB 182 LMSFDPG-FWSSSLKREILVTEGFSPLPLFSTTYKRAIDAR-----RKVAEALT 232
QY 232 DIKEBDQEEVEKYTDEDEMSKSDHVRKQRTDDLLGVWLKHSNLSSTQIIDLIIISLEFA 311
DB 233 VYVMKRREEEGAE-----RKMDLALLAADGSDDEIVFVALLVA 278
QY 312 GHETSSVAIALAIFFLQACPKAVEELREHELEIARAKKELGSELMDDYKMDTQCYI 371
DB 279 GETTSTMTLAVKFLTEPLALADLKEHEKIRAMKSD--SYLSWMSYKSMPTQCYV 336
QY 372 NETLGNVVRFLHRRALKDVKYKGYDIPSGMKVLPVISAVALDMSRYQPLFPMRWQ 431
DB 337 NETLVANLIGVFRAMTDVEIKGKIPKGMKVSFRFVADHDPARTFPMRWQ 396
QY 432 QONNGASSSGSFSFTWGN--YMPFGGPRLCAGSELAKLEMAVFIHLVLKFNNELEAD 490
DB 397 -----SNSVTGFSNVFTPPGGPRLCQGYELARVALSVLHRLVTFGFWVPAEQ 446
QY 491 DQFAPFPVDFPNGLPIRYSR 511
DB 447 DKLVEFPPTRTQKRYPIVKR 467

RESULT 5
AAG45021
ID AAG45021 standard; Protein: 491 AA.
XX AAG45021;
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56468.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-012180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

OY 491 DQFAFPYDFPENGPIRYSR 511
 1- 11 11 11 11
 Db 466 DKLVFPPTTRQKRYPIFYKR 486

RESULT 6
 AAG44570
 ID AAG44570 standard; Protein: 492 AA.
 AAG44570;
 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 55846.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 hydridisation assay; genetic mapping; gene expression control; promoter;
 termination sequence.
 KW
 OS Arabidopsis thaliana.
 PN EPI033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139494.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139762.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144002.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144864.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161982.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 36.5%; Score 978.5; DB 21; Length 492;
 Best Local Similarity 41.9%; Pred. No. 7e-78;

Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

QY 12 LLLPSSLLFLILKRRNRKTRFNLPPGKSGWPLGETIGYLRKYTAITTDGDFMOQH 71
 DB 27 LLLSSIAAGFL--LLRRTYRRMGLPPGSLGLPLIGFTFOLIAKTEENPEPFI DER 83
 QY 72 VSKYKRIYRNLTGCEPTIYASADAGLNRFILQNGRGLFECYPSISIGILGKWSMLVYGD 131

DB 84 VARYGVFWTHLGEPTTFISADPETNRFVLQNEGKLFECYSPASICNLLGKHSLLNKG 143
 QY 132 MHRDRSISLNFSLHARLRTILKDVRRHTFLVDSMOONSIFSAODEAKFTFNLMAKH 191
 DB 144 LHKRMSLJMSFANSSIHKHMLMDIDRLVRFNIDSSRVL--MEAKITTELIVKQ 201
 QY 192 IMSMDPEEETEOLKKEYVTFMKGVASAPNLPGTAVYHAKLQSRATTLKFTERRMEERKL 251
 DB 202 LMSFDPG-EMSESLRKREYLVLEGFFSLPLPLFSTYRKALQAR-----RKVAEALT 252
 QY 252 DIKEDEDEEEVKTEDAEKSKSDHVKORPTDDLLGWLKHSULSTEOIIDLILSLFLA 311
 DB 253 VVVKRRREEEGAE-----RKDKMLAALLAADDFSDSEIYDFVALLVA 298
 QY 312 GHETSSVAIALAIFLOACPAVELEBHEHEIARAKEGESEBLNMDYKKMFTOCVI 371
 DB 299 GYETTSITMLTAVKFLIETPLALNQLKEHEKRIAMSD--SYLEMSDYKSMFTQCIV 356
 QY 372 NETLRGNVYRFLHRAKLDKVRKGYDIPSGKYLPIYSAVHLDNSRYDQNLFPNRMQ 431
 DB 357 NETLRVANIIGVRRAMTDEIKYKIPKGMKVSSPRAVHLDPNHFKAARTFNPMRMQ 416
 QY 432 QQNGGASSSGSSTYGN--YMPFGGPRLCAGSELAKLEMAVFIHHLVLFKFWELAED 490
 DB 417 -----SNSVTGPSNVTPEFGGPRLCPEYELARVALSYELHRLVYGFMSVPAEQ 466
 QY 491 DQPPAPFVDPNGLPIRVS 511
 DB 467 DKLVFFPTTRKRPPIFVKR 487

RESULT 7
 AAG44572
 ID AAG44572 standard; Protein; 444 AA.
 AC AAG44572;

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55848.
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 OS Arabidopsis thaliana.

PN EPI033405-A2.
 PD 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 PF 25-FEB-1999; 99US-0121825.
 XX 05-MAR-1999; 99US-0123180.
 PR 05-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 25-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.

[illegible]


```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

```

```

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.18; Score 726; DB 21; Length 461;
Best Local Similarity 32.34; Pred. No. 1.5e-55;
Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;

```

```

QY 12 LLLPSLLSLFLILKRRN--RKRPNLPKGSMPFGERTGKLPYATITLGPMD 69
DB 7 LLIVSLCSALL-----RNMQRRTKNGLPPTGMPITGETTEFLKQGP----NMR 55
QY 70 OHVSKYKIRYSNLFGEPTIVSADAGLNRFITONEGRLEFECSPRSIGILKMSMLVLY 129
DB 56 NQRLRYGSFFKSHLGGPTILSMDSYVNRITLKMSGIVPGYQSMILITGTCNMAVH 115
QY 130 GDMHRDRSISLNFLSHARLTILKDKVERHTLFLDSMOONSIFSADEAKKFTF--- 185
DB 116 GSHRLRGSLLSLISSTWMDHILPKVDHMRSYLDQNNELFVIDIDKKHMAFLSL 175
QY 186 -----NMAKHIMMDGGEETBOLKKEVYTFMGVVSAPNLNCTAYHRLQSRATILK 240
DB 176 TQIAGNLKRPV-----EEKTAFFKLVGTLSVPIDLTGTYRCGIOARNIDR 225
QY 241 FIERKMEERKLDKEEDQEEBEVTEDEAEKMSGDHVRKQRTDDLLGMYLKHNN---L 296
DB 226 LIRELMQERR-----DSGE-----TFIDMLGILKMKKGNRYPL 258
QY 297 STEQILDILSLFAGHETSSVALATAIFLOCPKAVEELREHLEIARAKELGESET 356
DB 259 TDEIRIQVYVYIISGVEYVSTSMALKYLHDHPKALQELRABHLAFREKRQ--DEPL 316
QY 357 NMWDYKKMDFTQCYINFTLRGNVRFHLRKALKDVKRYKGDIPSGKKVLPVISAHLDN 416
DB 317 GLEDVKSMTFTRAVITYETSRLATTVNGVLRKRTYDLEINGLIRKGRITYYTRREIN 376
QY 417 SRYPDLFNPWRMOONNGASSSGSFSFNGNMYMPFGGPRICAGSELAKLEMAVFI 476
DB 377 NLVEDPLIFNPWRMKMSLEQ-----NSCFVFGGTYLCKCEKELGIYEISFL 425
QY 477 HHVLKFNWELAEDDQPAPEYDFPNGLPIRVS 510
DB 426 HYFTVRYRWEIEGDDELAVFPVPAPKGFHLRIS 459

RESULT 12
AAG46489
ID AAG46489 standard; Protein; 462 AA.
XX
AC AAG46489;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58493.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

```

OS Arabidopsis thaliana.
XX
PM EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148177.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153756.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148177.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149179.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155589.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 27.1%; Score 726; DB 21; length 465;
Best Local Similarity 32.3%; Pred. No. 1.5e-55;
Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;
OY 12 LLLPSSLILFLILKRRN--RKRPNLPKSGKGFPLGERTICYLKPYYATTTLGDPMQ 69

Db 11 LITIVLSCALL-----RNOMRYTKNGCLPPTMGWLPFGTEETBELKQGP-----NEMR 59
QY 70 OHVSKYKTYRNSLWFEPTIVSDAGLNREILONEGRLEFCSTYPRSIGTIGRWMLVLY 129
Db 60 NQRLRGSEFKSHLLGCPPLISMDSFVNRILNENSGVLPGVPGSLMDLILGTCNMAVH 119
QY 130 GDMHRRMRSISLNFSLHARLRTLLKDOVERHTLFVLDWQOONSIFSAODEAKFTF--- 185
Db 120 GSHRRLMRGSLISLSTMMRDLHPKVDHFMRSYLDQNMELVVIDQDKTKHAFSL 179
QY 186 -----NLMKHMISMDPGEETOLKKEYTFKGVVASPLNPGRYHIALSRATILK 240
Db 180 TQIAGMLRKPFV-----EEFTAFKPLVGLVSPIDLPNTVRCGIGQARNIDR 229
QY 241 FIERKKEERKLDIKEEDQEEVEKTEDEAMSKSDVHRKQRTDDLLGWLKHSN---L 286
Db 230 LRLMLQERR-----DSGR-----TTMDLGYLMKKEGNRYPL 262
QY 297 STFOILDLISLIPAGHETSSVAIALAIFPLQACPAVEELREHLEIARAKKELGSEL 356
Db 263 TDEIRDOVYTLISGYEYVSTSMALKYLDHPKALQDLRAEHILAFRRKQ--DEPL 320
QY 357 NMDDYKMKDFTQCVINETRLGNAVRFILRKALKDVRKYDYDIPSGKVLVPSAVHLDN 416
Db 321 GLEDVSMKFTFRAVIYETSRLATIVGVLRKTRDLEINCYLLPKGRIYVYTRREINYDA 380
QY 417 SRYDOPLFPMPOOONNASSSGSFGSTGNNYMPFGGPRLGCSELAKLEMAVFI 476
Db 381 NLTEDDLIFPMRMMKSLSEQ-----NSCFVRGGGRLCPGKELGIVEISSFL 429
QY 477 HHLVLFKNMELADDPFAFPVDFPNCGLPIRVS 510
Db 430 HYFVTRYRMEELGDELWFPVRFAPKGFHLRIS 463

RESULT 14
AAG11836
ID AAG11836 standard; Protein; 461 AA.
XX
AC AAG11836;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10708.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132483.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.

PR	21-OCT-1999	99US-0160741
PR	21-OCT-1999	99US-0160767
PR	21-OCT-1999	99US-0160768
PR	21-OCT-1999	99US-0160770
PR	21-OCT-1999	99US-0160814
PR	21-OCT-1999	99US-0160815
PR	21-OCT-1999	99US-0160880
PR	22-OCT-1999	99US-0160981
PR	22-OCT-1999	99US-0160989
PR	22-OCT-1999	99US-0161004
PR	23-OCT-1999	99US-0161405
PR	23-OCT-1999	99US-0161406
PR	23-OCT-1999	99US-0161407
PR	26-OCT-1999	99US-0161359
PR	26-OCT-1999	99US-0161360
PR	26-OCT-1999	99US-0161361
PR	28-OCT-1999	99US-0161920
PR	28-OCT-1999	99US-0161921
PR	28-OCT-1999	99US-0161923
PR	28-OCT-1999	99US-0162142

Query Match	26.6%;
Best Local Similarity	32.5%;
Matches 164; Conservative	9

QY	12	LLLPSSLSTLLFLFLILKRRN--KTRNRPNGSGMPPFLGRTIGYLAKPRTATTLGDFMQ	69
Db	1	MMILGLLIVYLCCTALLRMNQMRYSKKGPPGTMGPPIGETETFEFLKQGP----	DFMK 55
QY	70	QHVSVKGYKRSNLFGEPTIYSADAGNRFLLQNEGRLFECSYPRSRIGGILGKMSMLV	129
Db	56	NQRLRYGFFFSHILGCTTIVSMDELNRYILMNESSGLVAGTFQSMIDLITGTCNIAVH	115
QY	130	GDHMRDMSISLNFLSHARLTFTLLDKVERHTLFLVDSWOONSIFSADAEKRTFENLMA	189
Db	116	GPSHRLMGSLSLISPTMMKHLLPKIDPMRYNYLGGMDLLETVDIQEKTKNHAF--LSS	174
QY	190	KHMSMDGESEETEQLKKEYITFMKGVVASAPLNPATYAKHALQASRTATLKFLERKHEER	249
Db	175	LLOIAETFLKREVEVEYRTEFEFKLVAGTSLVPIDIPGTNYSGVQARNNIDRLLTLMQER	234
QY	250	KLDLKEEDQEEEEKTEDEAMSKSDHVRKQRTDDLLGLVLYKHSN---LSTEQILDLI	305
Db	235	K-----EGEGFTT-----DMLGYLMMKEDRNYLTIDKEIINDQV	267
QY	306	LSLTFAGHETSSVAIALAIFLQACPRAVEBELRENEHELIRAKKELEGESELMNDKKMD	365
Db	268	VTIIYSGEVETSTYSMAKLKYLHDHPALBELREHNLAIERKRP--DEPLITDDIKSMK	325
QY	366	FTQCYINETTLKLVNVRFLRKALKDVRKYCYDIPSCMKVLPVISAHLNDSRYDQNLF	425
Db	326	FTTRAVIETFSRLAIVNGVLRKTHDELNLQYLLPKMKRIYIVYRREINYPDTSLEYEDPMIF	385
QY	426	NPWRQOONNGASSSGSSFTWGNNTNMPGGGPRLCAGSELKLEMAVPIHLHLYLKFNW	485
Db	386	NPWRMESSLESKS-----YFLLEGGGVRLCPKGLGISTEVSFLHAFVYKRYW	434
QY	486	ELAEDDQFAEPFVDFPNGLEPIRYS	510
Db	435	EENEDKLMVFPVRSAPKGYHLKCS	459
RESULT 15			
AAG11835			
ID AAG11835 standard; Protein: 462 AA.			
XX	AC	AAG11835;	
XX	DT	17-OCT-2000 (first entry)	
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 10707.	
XX	FW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter;	

KW termination sequence.
 XX Arabidopsis thaliana.
 XX EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 XX 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123380.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139753.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149375.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151338.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.

```
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0133758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157533.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 26.6%; Score 712; DB 21; Length 462;
Best Local Similarity 32.5%; Pred. No. 2,5e-54;
Matches 164; Conservative 93; Mismatches 196; Indels 52; Gaps 8;

```
OY 12 LLLPSLSLLFLILKRRN--KTRFNLPPGKSGWPFLEGTIGYLRKYTTATLGDPMQ 69
DB 2 MMILGLVLIIVLCYALLMNMQRYSKGLPPTGMPPIFGETTEFLKQGP----DPMK 56
OY 70 QHVSXGKYYRNSLFGPEPTIVASDAGLNRFILQNEGRLEFCSPRSIGIIGKWSMLVLY 129
DB 57 NQRLKXGSEFKSHIIGCPITVSMDBELNRYITLINESKGLVAGYPOSMDILDTCTNIAVH 116
OY 130 GDMHDMRSISLNFSLSHARLITLLKDVERRHLFVLDWSQONSIFSAODEAKKFTFNMA 189
DB 117 GSHRLMRSGSLSLISPTMMKDHLLPKIDFMRNYLTCGWDDELEVDIOEKTQMAF-LSS 175
OY 190 KHMMSDPEEETEOUKKEVYTFMKGVASAPLNLPGTAYHKALOSRAITLKFIERKMEER 249
DB 176 LQIAETLLKPEVEEYKTEFFKLVGLTSLVPIDPTNYSRGVQARNNIDRLTTEIMQER 235
OY 250 KLDKEEDDEEEVTEDEAEKSKSDHAKKORTDDLLGWLKHSN---LSTEQILDLI 305
DB 236 K-----ESGERT-----DMGLYMKKEDNRYLLTDKEIRQVY 268
OY 306 LSLFAGHETSSVAIALAIFFLQACPKAVEELREHLEIARAKKEGESELNWDYKKMD 365
DB 269 VTIIVSGYETVETGMALKYLHDPKALEELRREHLAIKRRKP--DEPLTLDDIKSMK 326
```

```
OY 366 FTQCVINETLRIGNVRELHRAKADQVRKKGVDIPSGWKVLPVISAHLNDRYDQPIUF 425
DB 327 FTTRAVIFETSRLATIVNGVLRKRTTDLBNGYLIPKGRIRIYVYTRREINYDTSLYEDPMIF 386
OY 426 NPMRMQOONNGASSGSGSFSTWGNMYPFGGCPRLCAGSELAKLEMAVFIHHLVLRKFMW 485
DB 387 NPMRMKESLESK-----YFLTFGGVALLCFQKELGISVSSFLHIFTYTKRM 435
OY 486 ELAEDQPFAPFPFVDFPNGLPTRVS 510
DB 436 EENGEDKLIVFPFVSAPKGYHLKCS 460
```

Search completed: March 30, 2003, 12:04:18
Job time : 69 secs


```

Db 7 LLLLSIAAGFL---LLLRTRYRMGLPPGSLGLPIGETFQILGAYKTENPEPFIDER 63
QY 72 VSKYKIRYNSLFGPTTVSADAGLNRFLLQNEGRLEFECSPYRSIGLIGKMSMLVAGD 131
Db 64 VARYSVFETHLFGPTTIFSADPEFNRFVLQNEGRLEFECSPYRSIGLIGKMSMLVAGD 123
QY 132 MHRDRSISLNLSHARLRTLLKDYERHTLFLVDSMOONSIFSADDEKKTFFNLMAKH 191
Db 124 LKRRHSLTMSFANSIIKDHMLDIDRLVRENLDMSWSSRVLL--MEBAKKTFFELTVYQ 181
QY 192 IMSMDGEEETQOLKEVEYTFMKGVYSAVLNPGTAYHKAOSRATILKFIERKMEERL 251
Db 182 LMSFDPG--EWSLSKREYLVIEGFSPLPLFSTTYKKAIDAR-----RKVAEALT 232
QY 252 DIKEEDQEEVEYKTEDEAEMSKSDHVRKQRTDDLLGWVLKHSNSTEQIIDLISLFA 311
Db 233 VVMMRREEBEGAE-----RKMDLAALLAADGFSDEIYDFVALLVA 278
QY 312 GHETSVAIALAIFLQACPKAVEELREHLEIAAKKELGSELNWDYKMDTQCVI 371
Db 279 GYETSTMTTAVKFLTETPLALQKHEHEKIRAKMSD--SYLSWSDYKSMPTQCV 336
QY 372 NETLRIGVNVRLHAKKALDYKGYDIPSGKVLPIVSAVHLDNSRYDQPLFPMRMO 431
Db 337 NETLVANIGGVFRAMTDVEIKGKIPKGVSSPFAVHLDNHFADATFPMRMO 396
QY 432 QONNGASSSGSFSSTWGN--YMPGGGPRLCAGSELAKLENAVTHLVLKFNMLAED 490
Db 397 -----SNSVTTGSNVFTPPGGPRLCPGYELARVALSVLHRLVTFGSVPAEQ 446
QY 491 DQFAFPVDFPENGPIRYSR 511
Db 447 DKLVEPTTRQKRYPIYKR 467

RESULT 2
US-08-622-166A-4
Sequence 4, Application US/08622166A
Patent No. 5952545
GENERAL INFORMATION:
APPLICANT: KONCZ, CSABA
APPLICANT: MATUR, JAIDEEP
APPLICANT: SZEKERES, MIKLOS
APPLICANT: ALTMANN, THOMAS
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING CYTOCHROME
TITLE OF INVENTION: P450-TYPE PROTEINS INVOLVED IN THE BRASSINOSTEROID SYNTHESIS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Birch, Stewart, Kolasch & Birch, LLP
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/622,166A
FILING DATE: 27-MAR-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R.
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0147-0153P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 472 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-622-166A-4

Query Match 36.5%; Score 978.5; DB 2; Length 472;
Best Local Similarity 41.9%; Pred. No. 5e-89;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

12 LLLLSLILFLILKRRNRKTRPNLPKSGMPLGETGYLKPTATILGSPMOOH 71
7 LLLLSIAAGFL---LLLRTRYRMGLPPGSLGLPIGETFQILGAYKTENPEPFIDER 63
QY 72 VSKYKIRYNSLFGPTTVSADAGLNRFLLQNEGRLEFECSPYRSIGLIGKMSMLVAGD 131
Db 64 VARYSVFETHLFGPTTIFSADPEFNRFVLQNEGRLEFECSPYRSIGLIGKMSMLVAGD 123
QY 132 MHRDRSISLNLSHARLRTLLKDYERHTLFLVDSMOONSIFSADDEKKTFFNLMAKH 191
Db 124 LKRRHSLTMSFANSIIKDHMLDIDRLVRENLDMSWSSRVLL--MEBAKKTFFELTVYQ 181
QY 192 IMSMDGEEETQOLKEVEYTFMKGVYSAVLNPGTAYHKAOSRATILKFIERKMEERL 251
Db 182 LMSFDPG--EWSLSKREYLVIEGFSPLPLFSTTYKKAIDAR-----RKVAEALT 232
QY 252 DIKEEDQEEVEYKTEDEAEMSKSDHVRKQRTDDLLGWVLKHSNSTEQIIDLISLFA 311
Db 233 VVMMRREEBEGAE-----RKMDLAALLAADGFSDEIYDFVALLVA 278
QY 312 GHETSVAIALAIFLQACPKAVEELREHLEIAAKKELGSELNWDYKMDTQCVI 371
Db 279 GYETSTMTTAVKFLTETPLALQKHEHEKIRAKMSD--SYLSWSDYKSMPTQCV 336
QY 372 NETLRIGVNVRLHAKKALDYKGYDIPSGKVLPIVSAVHLDNSRYDQPLFPMRMO 431
Db 337 NETLVANIGGVFRAMTDVEIKGKIPKGVSSPFAVHLDNHFADATFPMRMO 396
QY 432 QONNGASSSGSFSSTWGN--YMPGGGPRLCAGSELAKLENAVTHLVLKFNMLAED 490
Db 397 -----SNSVTTGSNVFTPPGGPRLCPGYELARVALSVLHRLVTFGSVPAEQ 446
QY 491 DQFAFPVDFPENGPIRYSR 511
Db 447 DKLVEPTTRQKRYPIYKR 467

RESULT 3
US-08-724-466B-2
Sequence 2, Application US/08724466B
Patent No. 6063606
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: Blake, Cassels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
ZIP: M5L 1A9
COUNTRY: Canada
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,466B
FILING DATE: October 1, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996

```

```

ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-724-466b-2

Query Match      18.8%; Score 503; DB 3; Length 492;
Best Local Similarity 27.6%; Pred. No. 1,5e-41;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

QY 7 HTLLPLLLPSLLSLFL-----ILKRRNRKTRFNPFGKSGWPLFGTIGYLPK 58
D 4 YTLAVTFCTVLPVLLFLAALKMELMIRVDPNCKSPPLPGTWGLPFGTIGTQL-- 61
QY 59 YTATTLGDFMOQHYSKYKTRNSLFGEPITVSADAGLNRIILONEGLFECSTPRSTIG 118
D 62 ---LQRRKFLMKRKQKCYIKTHLFGNPTVYMGADNVROIILGHNKLVSVOMPASVRT 118
QY 119 ILGKMSMLVLVGDMDNR-----MRSISLNFSLHARLFTLLKDYVRHTLFLVDSW-OONS 172
D 119 ILGSDTLNSVNGVQHKKKRAIMRAFSDALEH-----YIPVIOQEVKSAIQEWLQKDS 172
QY 173 IFSADAEKKTFFNLMAKHIMSDPGEET--EQLKKEYVTFMGVSAPIPLPGTAYHK 230
D 173 CVLVYPEKKLMFRIAMHLLGFEPEQIKTDEQLVEAFEEIKMLFSLPIDVPSGLYR 232
QY 231 ALQSRATILKFERKMERKIDIKEEDOESEVKTEDAENKSDHYRKQRTDDLLGV 290
D 233 GLRRAR---NFHSKIEENIRKKTODDNEQKYKALQLL---IENSRSDE----- 279
QY 291 LKHSNLSLEQILDILSLFAGHETSSVAIALAIFFLQACPKAVEELREHLEIARAKE 350
D 280 ----PFSLOAKKEATLLEFGHETSTASTSLVFLGLNTEVQKVEE---VOEKVE 331
QY 351 LG---ESELNWDYKKMDFTQCVINETLRIGNVVRFLHRAKLDVRYKGYDIPSGWVL 406
D 332 MGMTTPGKGLSMELLDQKYGCVIKETLRINPPVPGGFRAVAKTFELNGYQIPKGMVYI 391
QY 407 PVISAVHLNDSRYDQPNLFNPRWQOONNGASSSGSFTWGNMYMPFGGPRICAGSE 466
D 392 YSICDTHVADVFPNKEEFQPERFW-----SKGLEDSRF--NYIPFGGGSRCVCYKE 442
QY 467 LAKLEMAVFIIHLVLFKFWELAEDDQ-----PFAFPVDFP 502
D 443 FAKVLKIFLVELTQHCWMLISNGPPYMKTGFTIYVDNLP 483

RESULT 4
US-08-882-164D-2
Sequence 2, Application US/08882164D
Patent No. 6306624
GENERAL INFORMATION:
APPLICANT: Petkovich, P. Martin, White, Jay A.,
APPLICANT: Beckett, Barbara R., Jones, Glenville
TITLE OF INVENTION: Retinoid Metabolizing Protein
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Blake, Casseels & Graydon
STREET: Box 25, Commerce Court West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5L 1A9
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage
COMPUTER: COMPAQ, IBM PC compatible
OPERATING SYSTEM: MS-DOS 5.1
SOFTWARE: WORD PERFECT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,164D
FILING DATE: June 25, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/667,546
FILING DATE: June 21, 1996
APPLICATION NUMBER: 08/724,466
FILING DATE: October 1, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Hunt, John C.
REGISTRATION NUMBER: 36,424
REFERENCE/DOCKET NUMBER: 50767/00010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 863-4344
TELEFAX: (416) 863-2653
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-882-164D-2

Query Match      18.8%; Score 503; DB 4; Length 492;
Best Local Similarity 27.6%; Pred. No. 1,5e-41;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

QY 7 HTLLPLLLPSLLSLFL-----ILKRRNRKTRFNPFGKSGWPLFGTIGYLPK 58
D 4 YTLAVTFCTVLPVLLFLAALKMELMIRVDPNCKSPPLPGTWGLPFGTIGTQL-- 61
QY 59 YTATTLGDFMOQHYSKYKTRNSLFGEPITVSADAGLNRIILONEGLFECSTPRSTIG 118
D 62 ---LQRRKFLMKRKQKCYIKTHLFGNPTVYMGADNVROIILGHNKLVSVOMPASVRT 118
QY 231 ALQSRATILKFERKMERKIDIKEEDOESEVKTEDAENKSDHYRKQRTDDLLGV 290
D 233 GLRRAR---NFHSKIEENIRKKTODDNEQKYKALQLL---IENSRSDE----- 279
QY 291 LKHSNLSLEQILDILSLFAGHETSSVAIALAIFFLQACPKAVEELREHLEIARAKE 350
D 280 ----PFSLOAKKEATLLEFGHETSTASTSLVFLGLNTEVQKVEE---VOEKVE 331
QY 351 LG---ESELNWDYKKMDFTQCVINETLRIGNVVRFLHRAKLDVRYKGYDIPSGWVL 406
D 332 MGMTTPGKGLSMELLDQKYGCVIKETLRINPPVPGGFRAVAKTFELNGYQIPKGMVYI 391
QY 407 PVISAVHLNDSRYDQPNLFNPRWQOONNGASSSGSFTWGNMYMPFGGPRICAGSE 466
D 392 YSICDTHVADVFPNKEEFQPERFW-----SKGLEDSRF--NYIPFGGGSRCVCYKE 442
QY 467 LAKLEMAVFIIHLVLFKFWELAEDDQ-----PFAFPVDFP 502
D 443 FAKVLKIFLVELTQHCWMLISNGPPYMKTGFTIYVDNLP 483

RESULT 5
US-08-882-164D-32
Sequence 32, Application US/08882164D
Patent No. 6306624
GENERAL INFORMATION:

```



```

Db 286 DMO-----ALKOS--STE-----LLEGGHETASATSLITLYGLYPHVLOKREELK 331
QY 343 EIAKAKKEGESELMNDYKMDFTQCVINETLRIGNVYRFLHRKALDKVARKGDIPSG 402
Db 332 SKGLLCKSNODNKIDMEILEQKTYGCVIKETLRINLPVPGGFVALKTEELNGYQIPKG 391
QY 403 WKVLPYISAVHLDNSRKYDOPNLFNPRWQOONNGASSSGSSTWGNVMPFGGPRLC 462
Db 392 WNVYISICDTHVAEIFTNKEEFNDRFSAPHPEDASRFS-----FIPFGGGLRSC 442
QY 463 AGSEIAKLEMAVEFHHVLKFNWELAE-----DDQPAFPFVDFP 502
Db 443 VKKEFAKILKIFVELARHCDMQLNGPPTKTSPTVYVNDLP 487

RESULT 7
US-08-882-164D-4
; Sequence 4, Application US/08882164D
; Patent No. 6306624
; GENERAL INFORMATION:
; APPLICANT: Petkovich, P. Martin, White, Jay A.,
; APPLICANT: Beckett, Barbara R., Jones, Glenville
; TITLE OF INVENTION: Retinoid Metabolizing Protein
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Blake, Cassels & Graydon
; STREET: Box 25, Commerce Court West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5L 1A9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: diskette, 3 1/2 inch, 1.4 Mb storage
; COMPUTER: COMPAQ, IBM PC compatible
; OPERATING SYSTEM: MS-DOS 5.1
; SOFTWARE: WORD PERFECT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/882,164D
; FILING DATE: June 25, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/667,546
; FILING DATE: June 21, 1996
; APPLICATION NUMBER: 08/724,466
; FILING DATE: October 1, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunt, John C.
; REGISTRATION NUMBER: 36,424
; REFERENCE/DOCKET NUMBER: 50767/00010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 863-4344
; TELEFAX: (416) 863-2653
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 497 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-882-164D-4

Query Match 16.4%; Score 439; DB 4; Length 497;
Best Local Similarity 27.0%; Pred. No. 3.7e-35;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;
QY 10 LPILLPSL-----LSLILPILK-----RRNKRTRNPPGSGWPELGETGYLK 57
Db 3 LPILLASLCTFVLPFLFLAATKLDLYCVSGDRSCALPDPGTMGPFEGETL-----58
QY 58 PYATATLGDPMOQHVSKRYGKTYRNSLFGSEPTIYSADAGLNRFILQNEGRLEBESYRSTG 117
Db 59 -QWVLGRKRLQKMKRRKRYGTYKTHLFGRTYAVYMGADVNRKLLDDRLVSYVHWASVRA 117
QY 118 GILGKSMALVYVGDMDRMSISLNFLSHARL--TILLKDVERTLFLVDSWQO--NS 172

```

```

Db 118 TILSGGCSNLDHSHKORRKYIMAFSREALCEYVPVITEEVGS-----LEQWLSCGER 173
QY 173 IFSADAEKKEFTFNMAKHMSDP-----GEEETEOUKREYVFMGVASAPLNLGTA 227
Db 174 GLVYPEYKRLMFRJAMNLLIGCEPQLAGDDBS-QQLVAFEEKTRNLSLPVDVPSG 232
QY 228 YHKALQSATLTKTERMEERKLDKEED-----QEEVEYKTEDEAMSKSDHVRKQRT 282
Db 233 LYRGKAKRNLHARIEQIRAKICGLRASEGQCKDALQLILHBSWE-----RERL 285
QY 283 DDDLGLWVAKNSNTEQIDLLSLFAGHETSVAIALAIFLQCPRAVEERREHL 342
Db 286 DMO-----ALKOS--STE-----LLEGGHETASATSLITLYGLYPHVLOKREELK 331
QY 343 EIAKAKKEGESELMNDYKMDFTQCVINETLRIGNVYRFLHRKALDKVARKGDIPSG 402
Db 332 SKGLLCKSNODNKIDMEILEQKTYGCVIKETLRINLPVPGGFVALKTEELNGYQIPKG 391
QY 403 WKVLPYISAVHLDNSRKYDOPNLFNPRWQOONNGASSSGSSTWGNVMPFGGPRLC 462
Db 392 WNVYISICDTHVAEIFTNKEEFNDRFSAPHPEDASRFS-----FIPFGGGLRSC 442
QY 463 AGSEIAKLEMAVEFHHVLKFNWELAE-----DDQPAFPFVDFP 502
Db 443 VKKEFAKILKIFVELARHCDMQLNGPPTKTSPTVYVNDLP 487

RESULT 8
US-09-615-192A-405
; Sequence 405, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003CAU
; CURRENT APPLICATION NUMBER: US/09/615,192A
; FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 405
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
; US-09-615-192A-405

Query Match 11.3%; Score 302.5; DB 4; Length 529;
Best Local Similarity 21.9%; Pred. No. 1.7e-21;
Matches 117; Conservative 104; Mismatches 204; Indels 109; Gaps 18;
QY 11 PILLPSLILFLILKRRNRKTRFNLPFGKSGMPPIGETIYIKPYATATLGDPMQO 70
Db 22 PMLLSLVPLLLFGLVLRKLRKPEF--PPGPRGIPVIGNML-----NMSELTRH 70
QY 71 HVS-----KXGKIYRNSLFEPTIYSADAGLNRFILQNEGRLE-----GYPYSI 116
Db 71 GLASLAKYGGIFHLRMGFLHMVAVSSPVARQVQLVHDGIFSNNRPAIATISLYLVDAD 130
QY 117 GGIIGKSMALVYVGDMDRMSIS--LNFLSHARLFTILKDKVERTLFLVDSWQONSIFS 175
Db 131 -----MAFAHYGPFWROMKLCVWKLFSSRRK-----ASBWE-----S 162
QY 176 ADEAK-----KFTFNMAKHMSMDPGEETEQKREYVTKK-----214
Db 163 VADVDVTVAVAGSEGRVAVNIGELVFELTRDIITRAAFGTSSTEG--ODEFTISLOFSK 221

```

QY 215 --GVVSAPINLEGTAY-----HKALOSRATITLFIENKME---ERKIDIEEDOESEE 262
DB 222 LFGAENIADPIFYLYLWIDPOGLTARLVKARQSIDGFIHIDIDHMDKKNKSSGGGDD 281
QY 263 VKTE--DEAEMSKSDHVRKQRTDDLLGWLKHS--NLSTEOILDLILSTLFGHETSSVA 319
DB 282 VTDWADLLATYSDEAKNESD-----LQNSIRLIDNIDKAILMDVFGTETVASA 335
QY 320 IALAIFFLOACPRAVEELREHLEIARAKKELGESLMDYDKMDPTQCVINETRLGN 379
DB 336 IEMAMELKRSPEDEKQVQGLADVGLDRVYES----DEKILYLKCCLEKTELRLHP 390
QY 380 VRFILHRKALKVRKKGIDIPSGKAVLPYISAVHLDNSYDDPNLFNPRMOOQNNGASS 439
DB 391 PIPLLHETAEADAVISGYRIPARSRYMINAMAIGRDPSGTEPDKFKPSRFL----- 443
QY 440 SSGSFSFTWGNVYFPGGPRLCAGSELAKLEMAVFIHILVLEKFMELAEDDP 493
DB 444 SGMPDYKGSNEFIPIFGSGRSCPGKQGLYALDMAVAHLLHCTWELPDGMP 497

RESULT 9

US-08-991-677-4
; Sequence 4, Application US/08991677A
; Patent No. 6252135
; GENERAL INFORMATION:
; APPLICANT: Chiang, Vincent L
; APPLICANT: Caraway, Daniel T
; APPLICANT: Smeltzer, Richard H
; TITLE OF INVENTION: Production of Syringyl Lignin in Gymnosperms
; FILE REFERENCE: 50617
; CURRENT APPLICATION NUMBER: US/08/991,677A
; CURRENT FILING DATE: 1997-12-16
; EARLIER APPLICATION NUMBER: US 60/033,381
; EARLIER FILING DATE: 1996-12-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Liquidambar styraciflua
US-08-991-677-4

Query Match 10.8%; Score 289; DB 4; Length 511;
Best Local Similarity 22.0%; Pred. No. 3.6e-20;
Matches 121; Conservative 95; Mismatches 196; Indels 138; Gaps 21;

QY 6 HTLLPL-LLLPSLLSLFLLLLRNRKTRFNLPRGKSGMPFIETIGYLKPTATT 64
DB 6 HEALDPLPMTLFTFTPLLLGLVSR--LRQLPYRPGKGLPVIGNML-MMDQLTHRG 62
QY 65 GDFMOOHVSKYKIRSNLFGPTIVSADAGLNRFILONEGRLEF-----CSYPR 114
DB 63 AKLANQ---YGGFLHKKMGFLHMAVSPDPAQVLOQDNIFSNRPATIASLVTYDR 118
QY 115 STIGILGKWSMLVYGDHNRKRSIS-LNFIASHARLRTLLKDVERTLFLVDSMOQNSI 173
DB 119 AD-----MAFAHGYPRWQRKLCVMKLFSSRRK-----AESWE--- 151
QY 174 FSAODE-----AKKFTFNLMKHMISMDGE-----ETE 203
DB 152 -SVROEVDASAVRNASNIGSTIVNIGELVAFALTKNITTYRAFGTISHEDQDEVAILOES 210
QY 204 QL-----KREYTFMKGV---VSADPLNPGTAYHKALQSRAITLFIERKMEERKLDIKE 255
DB 211 QLFQAFNLADFTPMUKWYPOGINVRLN-----KARGLDQFIKIIDH---IQK 257
QY 236 EDQEEBEVKT-----DEAEMSKSDHVRKQRTDDLLGWLKHSNLSSTEOILDLI 305
DB 238 GSKNESEVDYDMDLLAFYGEAEKAVSESDD-----LQNSIKILTKNIKAI 303
QY 306 LSLFLAGHETSSVALAIFLQACPRAVEELREHLEIARAKKELGESLMDYDKMD 365

DB 304 MDVFGGETVASALEMATELTKSPEDLKQVQGLAVVGLDRVREK-----DFEKL 358
QY 366 FTQCVINETRLGNVVRFLHRKALKDVRKKGIDIPSGMWLVPIASAVHLDNSRYDDPNLF 425
DB 359 YLKYVLKELRLHPRPILLLHETAEADVAEYGYIIPAKSKVMINACAIQDKRSMADPDTF 418
QY 426 NFRMOOQNNGASSSGSFSFTWGN--YMPFGGPRLCAGSELAKLEMAVFIHILVLEK 483
DB 419 RRSRFLK-----GVPDFK--GNNFEFIPIFGSGRSCPGKQGLYALETVAHLLHCF 469
QY 484 NMELEDDP 493
DB 470 TWELPDGMP 479

RESULT 10

US-08-948-564-6
; Sequence 6, Application US/08948564
; Patent No. 6121512
; GENERAL INFORMATION:
; APPLICANT: Siminszky, Balazs
; APPLICANT: Dewey, Ralph E.
; APPLICANT: Corbin, Frederick T.
; TITLE OF INVENTION: No. 6121512a1 Cytochrome P-450 Constructs and
; TITLE OF INVENTION: Methods of Producing Herbicide-Resistant Transgenic Plants
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 6121512th Carolina
; COUNTRY: USA
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/948,564
; FILING DATE:
; CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Virginia C.
REGISTRATION NUMBER: 37,092
REFERENCE/DOCKET NUMBER: 5051-409
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-854-1400
TELEFAX: 919-854-1401

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-948-564-6

Query Match 10.4%; Score 279.5; DB 3; Length 513;
Best Local Similarity 22.9%; Pred. No. 3.2e-19;
Matches 118; Conservative 82; Mismatches 216; Indels 99; Gaps 17;

QY 15 LPSLLSLFLLLLRNRKTRFNLPRGKSGMPFIETIGYV---KPYATTLGDFMOOH 71
DB 15 LAFISGLIF--FLKQSKSKKFNLPDGPWPYIGNLFQVARSKPF-----EYVDV 67
QY 72 VSKYKIRSNLFGPTIVSADAGLNRFILONEGRLEFECSTYR----- 114
DB 68 RLKYSIFTLKGTFTMTILIDAKLVHEAMIOKQATVATRPDENPTRTLFSENKFTVNA 127
QY 115 STIGILGKWSMLVYGDHNRKRSISLNFIASHARLRTLLKDVERTLFLV-----DSW 168
DB 128 TYGPV---WKSLL-----RRNVQNNLSSTRLKE--FRSVRDNADKILNRLKDEAE 173

QY 169 QONSIPODEAKFTFNMAKIHMSMDPEETBOLAKREYTPMKGVSA----- 219
 Db 174 KNGVWVLKDAFAVFCILVAMCGLEMEDETERIDQ---VMSVLTLPDPRIDYL 229
 QY 220 PUNL-P-GTAYHKALOSRAITLKEFERKMEBERLIDKEEDQEEBEVKTDEAMSKSDHY 277
 Db 230 PILSPFFSKORAKALEVRQVEFLVPIIQRRAIIONG-----SDHT 273
 QY 278 RKORTDDLL-----GWLKHSNLSTEQDILDLISLFAGHETSSVAIALAIFLOACP 331
 Db 274 ATTFSLIDTLFDLKVBS---KKSAPDAELVYLCSEFLNGCTPTTAFAVWGLAQLIANP 330
 QY 332 KAVEELREBEHLEIARAKKELGESELMWMDYKKWDFQCIVNETLRLGNVVRFLHRAKLD 391
 Db 331 MNOTKLYEE-----IKRTVGEKKVDEKVEKMPYLHAYVKEILRKHPFTHTHAYTE 384
 QY 392 -VAYKGYDIPSGMKVLPVISAHLNDSRYDQPLFNPWMOQONNASSSGSFSFTWGN 450
 Db 385 PTTLGSDIDIPIDANVEYTPALAEDEPKNMLNPKFDPERISGGEADITG---VTGV 439
 QY 451 NYMFGGPRLCAGSELAKLEMAVFIHHLVLEKFNW 485
 Db 440 KMPFVGGRICPGILAMATVHIHLMARWQEEFW 474

RESULT 11
 US-09-144-367-2
 ; Sequence 2, Application US/09144367
 ; Patent No. 6432639
 ; GENERAL INFORMATION:
 ; APPLICANT: Lichteir, Jay
 ; APPLICANT: Guido, Marco
 ; TITLE OF INVENTION: GENOTYPING OF HUMAN CYP3A4
 ; FILE REFERENCE: SEQ-12P
 ; CURRENT APPLICATION NUMBER: US/09/144,367
 ; CURRENT FILING DATE: 1998-08-31
 ; PRIOR APPLICATION NUMBER: 60/058,612
 ; PRIOR FILING DATE: 1997-09-10
 ; NUMBER OF SEQ ID NOS: 58
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 503
 ; TYPE: PRT
 ; ORGANISM: H. sapiens
 US-09-144-367-2

Query Match 10.2%; Score 273; DB 4; Length 503;
 Best Local Similarity 22.4%; Pred. No. 1.4e-18;
 Matches 120; Conservative 101; Mismatches 222; Indels 92; Gaps 21;
 QY 9 LPLLLPSSLILFLILKRRNRKTRFNLPGKSGWPLGETIGYLPYATL 64
 Db 3 LIPDLAMETWLLAVSLVLYLGTSHGFLKGLIGPPLPPLFLGNIILYHKGFCM--- 59
 QY 65 GDFMGOHVSKYGIYNSNLGEPPTVSADAGLNRFLIQLONGRLFECSYPSIGIGL-K 122
 Db 60 --FDMCHCKYKGYGFDQOQVLAITDPDMIKTVLVC--CYSVPTNRPPGPGVGMK 115
 QY 123 WSNLVLVGDMHDMRS-ISLNLFLSH-----ARKTILLADVENH-----TLFVL 165
 Db 116 SAISLIDEHEWKRSLSTFTSGKLKENVPILAIQGYDLVNRNLREAEATGKPVTL--- 172
 QY 166 DSMQONSIFSA--ODEAKKFTFNMAKIHMS-MDPEETEOLK--EYV-TFMKGVVASP 220
 Db 173 ----KDVFGAYSMVDYITTSFGVNIDSLNPPDPFVFNKTLIRPFLDPFLSTIVFP 227
 QY 221 LNPGLAYHKALOSRAITLKEFERKMEERLIDKEEDQEEBEVKTDEAMSKSDHYRKQ 280
 Db 228 FLIPLEIVLNICVPREVTNFL-RKSVKRMKESRLIEDTQKRV---DFLOL-MDSQNSK 282
 QY 281 RPDDDLGWLKHSNLSTEQDILDLISLFAGHETSSVAIALAIFLOACPRAVEELREE 340

Db 283 ETE-----SHKALSDLEVAOSIIFIFAGYETTSVSFLMYELATRPDQOKLOE 334
 QY 341 HLEIARAKKELGSELNMDYKKWDFQCIVNETLRLGNVVRFLHRAKLDVRYKYDIP 400
 Db 335 -IDAVLPNK-----APPTDYVLQMEYLDVYVNETLRLFLPAMLERVCKKNDVEINGWFI 389
 QY 401 SGKVLVISAHLNDSRYDQPLFNPWMOQONNASSSGSFSFTWGN-----YMPF 455
 Db 390 KGVWVMIPTSYALHRDPKYTWPEPEKFLPERVSKNRK-----DNIDPYITPF 435
 QY 456 GGGPRLCAGSELAKLEMAVFIHHLVLEKFNELADDPFAFPFDPNGSLPIRYS 510
 Db 436 GSGFRNCTGMRFLMNMKIALIRVLQNSFRPKCKETO-----IPLKLS 478

RESULT 12
 US-09-126-420A-18
 ; Sequence 18, Application US/09126420A
 ; Patent No. 6376753
 ; GENERAL INFORMATION:
 ; APPLICANT: BATAARD, YANNICK
 ; APPLICANT: ROBINEAU, TIBURCE
 ; APPLICANT: DURS, FRANCIS
 ; APPLICANT: MERCK-REICHART, DANIELE
 ; APPLICANT: DIDIERJEAN, LUC
 ; TITLE OF INVENTION: PURIFIED CYTOCHROME P450 CYP76B1 FROM HELIANTHUS
 ; TITLE OF INVENTION: TUBEROSUS AND ITS APPLICATIONS AS BIOCATALYST IN
 ; TITLE OF INVENTION: PARTICIPANTS AND FOR THE DEGRADATION OF ENVIRONMENTAL
 ; TITLE OF INVENTION: POLYMERIS AND FOR ALTERING THE RESISTANCE OF PLANTS S
 ; FILE REFERENCE: 03715.0032
 ; CURRENT APPLICATION NUMBER: US/09/126,420A
 ; CURRENT FILING DATE: 1998-07-30
 ; PRIOR APPLICATION NUMBER: 60/054,351
 ; PRIOR FILING DATE: 1997-07-31
 ; NUMBER OF SEQ ID NOS: 27
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 18
 ; LENGTH: 504
 ; TYPE: PRT
 ; ORGANISM: Solanum melongena
 US-09-126-420A-18

Query Match 10.1%; Score 271; DB 4; Length 504;
 Best Local Similarity 21.7%; Pred. No. 2.2e-18;
 Matches 111; Conservative 95; Mismatches 231; Indels 74; Gaps 17;
 QY 12 LPLLLPSSLILFLILKRRNRKTRFNLPGKSGWPLGETIGYLPYATL 70
 Db 13 IILIPA-----FLIFESQKNTTKSSYRPPGPGPLIFGMNFEIGTEPYKMAV---LRQ 63
 QY 71 HVSRYGIYNSNLGEPPTVSADAGLNRFLIQLONGRLFE-----ECSYPSIGIGI 119
 Db 64 ---KYGVVLMKESYTTMAYVOTQASSELKFNHIDISFANRVIVDVQASHYTGSLA-- 118
 QY 120 LKWSMLVLVGDHDMRSI-SLNLFLSHARLITLADVENHFLVLVDSMQONSIFSAOD 178
 Db 119 ----IAPGPFWRFORICTIEMFVHKIIS--TEPVARRKCYDNLKWIEMKANSABK 170
 QY 179 ----EAKKF--TFNMAKIHMSMDPEETBOLAKREYTPMK-----GVSAPLMDPG 225
 Db 171 GSGIEVTRFVFLASFNMLNLSKDLADESEASEFFIAMKINMNGSIANSDIIPP 230
 QY 226 TAYHKALOSRAITLKEFERKMEERLIDKEEDQEEBEVKTDEAMSKSDHYRKQRTDD 285
 Db 231 LKTFDLSLKKMARMDGKAVETMSMLKE--REEKKKGTEKKDLDVLILEQTGKD 288
 QY 286 LIGVNLKHSNLSTEQDILDLISLFAGHETSSVAIALAIFLOACPRAVEELREEHLEIA 345
 Db 289 ----EPAKISEHEIRIFVLEMFIACTETTSSEVALTELRLHPAMAKVKTEISQAI 342
 QY 346 RAKKEGESELMWMDYKKWDFQCIVNETLRLGNVVRFL-HRAKLVRYKGYDIPSGWK 404
 Db 343 EPNKRFDS-----DIENLPYMQAVIKESLRLHPLPFLIPRETIOPTKMGVDVPRDQ 397

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2003, 12:08:26 ; Search time 51 Seconds

(without alignments)
590.691 Million cell updates/sec

Title: US-09-502-426a-2

Perfect score: 2681
Sequence: 1 MPEDEHHTLLPLLLPSLTS.....FAFPVDFPGLPIRYSRLL 513

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppa/PC7_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppa/PC7US_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	292	10.9	516	9 US-10-067-534-3	Sequence 3, Appli
2	290.5	10.8	514	10 US-09-947-027-4	Sequence 4, Appli
3	290.5	10.8	514	12 US-10-091-009-4	Sequence 4, Appli
4	289	10.8	511	10 US-09-796-256a-4	Sequence 4, Appli
5	285.5	10.6	502	10 US-09-739-254-69	Sequence 69, Appl
6	285.5	10.6	502	10 US-09-904-615-69	Sequence 69, Appl
7	275	10.3	496	9 US-10-067-668-10	Sequence 10, Appl
8	275	10.3	496	10 US-09-945-301-4	Sequence 4, Appli
9	272	10.1	470	9 US-10-067-668-12	Sequence 12, Appl
10	256.5	9.6	520	10 US-09-992-901-2	Sequence 2, Appli
11	251	9.4	515	10 US-09-796-138-19	Sequence 19, Appl
12	251	9.4	515	10 US-09-909-903-19	Sequence 19, Appl
13	251	9.4	544	9 US-10-067-668-8	Sequence 8, Appli
14	249	9.3	503	10 US-09-796-138-18	Sequence 18, Appl
15	249	9.3	503	10 US-09-909-903-18	Sequence 18, Appl
16	249	9.3	503	10 US-09-957-997-3	Sequence 3, Appli
17	249	9.3	509	10 US-09-817-582-4	Sequence 4, Appli
18	247.5	9.2	493	9 US-10-103-520-2	Sequence 2, Appli
19	247.5	9.2	505	10 US-09-910-689-307	Sequence 307, App

20	247.5	9.2	505	12 US-10-010-742-307	Sequence 307, App
21	246	9.2	509	9 US-10-028-072-108	Sequence 108, App
22	246	9.2	509	9 US-10-121-049-108	Sequence 108, App
23	246	9.2	509	9 US-10-123-904-108	Sequence 108, App
24	246	9.2	509	9 US-10-140-470-108	Sequence 108, App
25	246	9.2	509	9 US-10-175-746-108	Sequence 108, App
26	246	9.2	509	9 US-10-176-918-108	Sequence 108, App
27	246	9.2	509	9 US-10-176-921-108	Sequence 108, App
28	246	9.2	509	9 US-10-137-865-108	Sequence 108, App
29	246	9.2	509	9 US-10-140-474-108	Sequence 108, App
30	246	9.2	509	9 US-10-142-431-108	Sequence 108, App
31	246	9.2	509	9 US-10-143-114-108	Sequence 108, App
32	246	9.2	509	9 US-10-140-002-108	Sequence 108, App
33	246	9.2	509	9 US-10-142-419-108	Sequence 108, App
34	246	9.2	509	9 US-10-123-262-108	Sequence 108, App
35	246	9.2	509	9 US-10-142-423-108	Sequence 108, App
36	246	9.2	509	9 US-10-121-050-108	Sequence 108, App
37	246	9.2	509	9 US-10-141-755-108	Sequence 108, App
38	245.5	9.2	508	9 US-09-925-299-905	Sequence 905, App
39	245.5	9.2	508	10 US-09-925-299-905	Sequence 905, App
40	244.5	9.1	524	9 US-09-992-598-264	Sequence 264, App
41	244.5	9.1	524	9 US-09-989-293A-264	Sequence 264, App
42	244.5	9.1	524	9 US-10-063-547-54	Sequence 54, Appl
43	244.5	9.1	524	9 US-09-989-735-264	Sequence 264, App
44	244.5	9.1	524	9 US-09-980-444-264	Sequence 264, App
45	244.5	9.1	524	9 US-09-989-730-264	Sequence 264, App

ALIGNMENTS

RESULT 1
US-10-067-534-3
Sequence 3, Application US/10067534
Publication No. US20020187538A1
GENERAL INFORMATION:
APPLICANT: Essenberg, Margaret K.
APPLICANT: Chen, Xiao-Ya
APPLICANT: Luo, Ping
TITLE OF INVENTION: cDNA Clone of (+)-Delta-Cadinene-8-Hydroxylase Gene from *Colt*
FILE REFERENCE: 006602-113
CURRENT FILING DATE: 2002-02-07
PRIOR APPLICATION NUMBER: US 60/267,160
PRIOR FILING DATE: 2001-02-07
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 516
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-10-067-534-3

Query Match 10.9%; Score 292; DB 9; Length 516;
Best Local Similarity 23.6%; Pred. No. 7.8e-16;
Matches 120; Conservative 92; Mismatches 220; Indels 76; Gaps 17;
12 LLLPSLLFLFLILKRRNRKTRFNLPKSGMFLGETIGYLKPYATTLGDMOQH 71
20 IYLLTVSEILWIF-----KSPQSLPDPGRGLPIV-NLPFLDPDLHTFANLAOSH 73
72 VSKRYKYSNLFGEFTVSADAGLNRFLLQNGRLF-----CSYPSIGIGLQKSM 125
74 ----GFIFKLNIGSKLTIYVNSPFLAREILKQDINFNRDVLNGRAATYGGIDIVMP 129
126 LVLVGEMHMDMSIS-LNLSHARLFT---ILKQVRETLFVLDSMOONSIFSADDEAK 181
130 ---YGAEMWOLKIKICVLKLSRTLDSEYELRKVYRETRLYVEGKQSPYKVDQL 186
182 KFTFNMAKHISMDDGEETDLKREYTFMKGVV-----APLNL 223
187 LTMNMLTMMNMGSGYKADMEVGTFE-----KGVSEITRLISEPHVSDFFPMLARFDL 242

```

QY 224 PGTAVHKAJOSRATILTKIERKMEERKADIKKEODEEEVTEDEAEKMSQHVAKQRTD 283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 243 QGLVVRMGVCAR-ELDANVLDRAIBEMK---PLRGDDDEVDFQLQMLKLD----QGED 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 284 DDLLGWLKHSNLSTEOILLDLISLPFGHETSSVALATLIFLOACKRAVEELREHLE 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 SEV-----PITFNHKALETDMVVGSTDTSTTTFIEFAAEELMSNPBLIKRAOEELDE 346
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 IARAKKEGESELNMDDVKKDETCQVNETRLRGVNVRF--HRKALKDVRKYGDIDPS 401
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 347 VVGKDNVIEESHIT-----RLPYLAIKKEITRLKPLPLVPHRPAENTV-VGCTYTTPK 400
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 GKKVLPIVISAVHLNDSRYDQNLFPNPMWMOOONNCASSGSGSEFTSWCNTMYMPEGGPRL 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 401 DTKIEVNVMSIORPNNVNMENPTEEFPERFIINN-----SCDFGANYSYEPFGSGARI 453
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 462 CAGSELAKLEMAVFTIHLVLPFNWELAE 489
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 454 CAGVALAERWVLYTLATILSHSDMKIPE 481
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 2
 US-09-947-027-4
 : Sequence 4, Application US/09947027
 : Patent No. US20020124281A1
 : GENERAL INFORMATION:
 : APPLICANT: Chiang, Vincent Lee C.
 : APPLICANT: Li, Laiyeng
 : TITLE OF INVENTION: METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND COMPOSITION
 : FILE REFERENCE: 066040-9718
 : CURRENT APPLICATION NUMBER: US/09/947,027
 : PRIOR FILING DATE: 2001-09-05
 : PRIOR APPLICATION NUMBER: 60/230,086
 : NUMBER OF SEQ ID NOS: 14
 : SOFTWARE: PatentIn version 3.0
 : SEQ ID NO 4
 : LENGTH: 514
 : TYPE: PRT
 : ORGANISM: aspen populus tremuloides
 : US-09-947-027-4

	Query Match	10.8%;	Score 290.5;	DB 10;	Length 514;	
	Best Local Similarity	22.6%;	Pred. No. 1e-15;			
	Matches 123;	Conservative	98;	Mismatches 230;	Indels 93;	Gaps 19;
Oy	11	PLLLPSLSLLELILIKRRNRKTRFLPEQSGWPLGETIGYLKPATVTLGDEMGO	70			
Db	12	PMSEFLIYSSIFFEGLSLRLR--RLYPGPCKLPVWG-SMHMMQITTHGKLAKLO	68			
Oy	71	HVSKGKIYRSNLGEPTIVSADALNFIINBGRLEE-----CSTPRISIGILL	120			
Db	69	----YGLIEHHRMCGYLAHWTVSSPEIAQLQVONDFINSRNPAITAISTLYTYRAD---	120			
Oy	121	GRWSMLVLVDGMHRDMRSIS-LNFESHAIRLT-----ILTKDVERH-----T	161			
Db	121	--MAFAHYGFPMWRMRKLCYVKLFESRKRAESWEVSVRDEYSMLKTVEANICKPYNLGEL	177			
Oy	162	LFLVDSMOOONSIFFSQODAKKFTEFLMLAKNHMSMDPGEEETEQLKKEYTVMKGVSAPL	221			
Db	178	IFTL--TMNITYTRAFGAK-----NEGODEFETIKIQEF-SKLFEGAFNMSD	219			
Oy	222	NLPGRVAHKALQSRTILIKTERKKWEERKLD-IKEEDOEVEEVTEDAEAMSKSD----	275			
Db	220	FIPMLGWIDPGLSARLVK--ARKLDDRFIDSITDHDIQRKQNKFSEDAEFDWDVDMIA	277			
Oy	276	---HVKRQTDDDLGAVWLKHSHNSLSTQCIDLISLTFAGHTSSVAITALAIFFLQACP	331			
Db	278	FYGEEARAYDESDDL-----QKAISLTKNIKAILINDVAFEGGTETASALEWMAMELMKSP	333			
Oy	332	KAVEELREHEIELIARAKKELGBSELNDWDYKMKDFTQCVINETFRLGNVYFRLLARRALKD	391			

Db	334	EDOKVOCLEIAVVGLEERVEESID-----KLFELCALKEITLRMPPIPLLHETSED	3688
Oy	392	VRYKGYDIPSGMKVLPYSAVHLNBSRYDOPLENPRKMOQNNAGSSGSGSFSTGNN	451
Db	369	AEVAGYFLFPKOTRWINAYAIGRKNSKNEDDAKPRFLK-----PCVPDEK--GNH	4399
Oy	452	--YMPGGGPRLCAGSELAKLEMAVFTHHLVLKENWELAEADDOPFAPRVDFPNCPIRY	509
Db	440	FEETPFSGGRSRSCGMQGLYTLDAVAHLHCFTWELPGMKPSELMDTMDGLTAPRA	499
Oy	510	SRIL 513	
Db	500	TRLV 503	

```

1      RESULT 3
2      US-10-091-009-4
3      : Sequence 4, Application US/10091009
4      : Patent No. US20020138870A1
5      : GENERAL INFORMATION:
6      : APPLICANT: Chiang, Vincent Lee C.
7      : APPLICANT: Li, Laiyeng
8      : TITLE OF INVENTION: METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND
9      : TITLE OF INVENTION: COMPOSITION, AND
10     : TITLE OF INVENTION: CELLULOSE CONTENT IN PLANTS
11     : FILE REFERENCE: 066040-9718
12     : CURRENT APPLICATION NUMBER: US/10/091,009
13     : PRIOR FILING DATE: 2002-03-06
14     : PRIOR APPLICATION NUMBER: 09/947, 027
15     : PRIOR FILING DATE: 2001-09-05
16     : PRIOR APPLICATION NUMBER: 60/2230, 086
17     : PRIOR FILING DATE: 2000-09-05
18     : NUMBER OF SEQ ID NOS: 14
19     : .SOFTWARE: PatentIn version 3.0
20     : SEQ ID NO 4
21     : LENGTH: 514
22     : TYPE: PRF
23     : ORGANISM: aspen populus tremuloides
24     : US-10-091-009-4

```

Query Match	10.8%	Score 290.5	DB 12	Length 514
Best Local Similarity	22.6%	Pred. No. 1e-15		
Matches 123	Conservative	98	Mismatches 230	Indels 93
				Gaps 19
QY	11	P L L L P S L L S L L F E I L L K R N R K T R E M L P R G S G M P L G E T T I G K R Y A T T A T T L G D F M Q	70	
Db	12	12 P M S L F L I V I S S L F F E G L L S R L R R - R L Y P P G K L P V G - S M H M D Q T T H R G L K L A K	68	
QY	71	H V S K G K I Y R S N L G E P P I V S A D A G I N F I L I O N E G R L F E - - - - - C S P R A I G T L	120	
Db	69	69 - - - - - Y G G L F H M R M G Y L H M W T V S S P E I A Q V L O V O N I T S N R P A N I A I S T I L Y I D R A D - - - -	120	
QY	121	G K W S M L V V G D M H R D M R S I S - I N F L S H A R L T - - - - - I L K D V E R H - - - - - T	161	
Db	121	121 - - - M A F A H Y G P F M R O R K L C V M K L F R S R K A B S W E S V R D E V S M L T V E A N I C K P V N L G E L	177	
QY	162	L E V I D S M Q O O N S I F F S Q O D A K K F T F I L M A K H I M S M P G E E T E T O L K R E Y T F P K G V S A P L	221	
Db	178	178 I F T L - T M N I T T Y R A F A G A - - - - - N E G O D E F T K I Q I O E R - S K I F G A F M S M D	219	
QY	222	N L P G T A Y H K A L Q S R A T I L K F T E R K M E E R K L D - I K E E Q O E E E V T E D B A E M S K S D - - - -	275	
Db	220	220 F I P M G I W I D P G L S R L V K - A R K A L D F I D S I D D H I Q K R O N K F S E D A E T D W D M L A	277	
QY	276	276 - - - - - H V K R O T D D D L I G V N L K H S N I S T E Q I L D L I L S L F A G H E T S S V A I A I A I F P L Q A C P	331	
Db	278	278 F Y G E A R K V E S D S D L - - - - - O K A I S L T K N I K A I I M D V M F G E T V A S A I E M V M A E L M K S P	333	
QY	332	K A V E L R E H E L I A R A K K E L G E S E L N M D Y K K M D T Q C V I N E T L L A G V N R V L R L H K A L K D	391	
Db	334	334 E D Q R V O Q E L A E V V G L E R R V E S D I D - - - - - K L F L K A K A L E T L M A P P I P L L H E T S D	388	

QY 392 VRYGYDIPSGWKVLPVISAHLNDRYDOPNLFNPMWMOONNGASSGSGSFSTWGN 451
DB 389 AEVAGYIFPKOTRYMINVAIGRDKNSWEDPAFKPSRLK-----PGVDFK--GNH 439
QY 452 --YMPFGGPRLCAGSELAKLEMAVFIHLVYKFMWELAEEDOPAFPFVFPNGLPTRV 509
DB 440 FEPIFGSGRRSGPQGLGYTLIDIAVAHLHCFWELPDGMRKPSLMDTMFGLTPRA 499
QY 510 SRIL 513
DB 500 TRIV 503

RESULT 4

US-09-796-256A-4
Sequence 4, Application US/09796256A
Patent No. US2002007847A1
GENERAL INFORMATION:
APPLICANT: Chiang, Vincent L
APPLICANT: Cattray, Daniel R
TITLE OF INVENTION: Production of Stryngyl Lignin in Gymnosperms
FILE REFERENCE: 50617/G-3532.0
CURRENT APPLICATION NUMBER: US/09/796,256A
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US60/033381
PRIOR FILING DATE: 1996-12-16
PRIOR APPLICATION NUMBER: 08/991677
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 4
LENGTH: 511
TYPE: PRT
ORGANISM: Liquidambar styraciflua
US-09-796-256A-4

Query Match
Best Local Similarity 10.8%; Score 289; DB 10; Length 511;
Matches 121; Conservative 95; Mismatches 196; Indels 138; Gaps 21;

QY 6 HHTLLPL-LILPSSLILFLILKRNKRTNLPKSGMPFLEGTGYLKPYATT 64
DB 6 HEALQPLPMTLFLILPILLLGLVSR--LNRRLPYPPGPKGLPIGNML-MMDOLTHGCL 62
QY 65 GDFMOHVSXGKIYNSNLFGEPTIVSADAGNRFILONEGRLEF-----CSYPR 114
DB 63 AKLAKO-----YGGFLHKLKMGFLMVAVSTPDMAQVLYODNIFSNRPATTAISYLYDR 118
QY 115 SIGGILGRKSMVLVGDMDRMRKRSIS-LNFISSHARLRTILKDYERHTLFVLDGMOONSI 173
DB 119 AD-----MAFAHYGPFWRQMRKLCYMKLSRRK-----AESWE----- 151
QY 174 FSAODE-----AKKFTENLMKHHMSMDPE-----EETE 203
DB 152 -SVRDEVDAAVAVASNGSTVNGELVAFALTKNITRYRAFGTISHDDODEFVALIIGFTS 210
QY 204 QI-----KKEYTTEKGV--VSAPLNLPGTAYKALQSAATILKTERKMEERKIDIRE 255
DB 211 QILGAFNIADIFPMKWPQGINVRN-----KANGALDGIIDKIDH---IQK 257
QY 236 EDDEEEVKT-----DEAKMSDVRKORTDDLLGWLKHSNLSSTBOILDLT 305
DB 258 GSKNSEEDVDMDLLATFGEELKVSDD-----LQNSIKLTKONIKAI 303
QY 306 LSLFAGHERSSVAIALAIFLQACRAVEELREHLEIARAKKELGSEELNMDYKMD 365
DB 304 MDVAFGGTEVVAIAIEMAMTELKSPEDLKQOELAVVAGIDRVERK-----DFEKL 358
QY 366 FTQCVINETLRIGNVVRFHLRAKLDVARKGIDIPSGWKVLPVISAHLNDRYDOPNLF 425
DB 359 YLKCIVKEVLRIRPPIPLLLHETIADAEVGGYIIPAKSRVMINACALGRKNSMDPDR 418

QY 426 NPMWMOONNGASSGSGSFSTWGN--YMPFGGPRLCAGSELAKLEMAVFIHLVYK 483
DB 419 RSRRLK-----DGVDFK--GNNEFIIPSGRRSGPQGLGYLTALETVVAHLHCF 469
QY 484 NWELEDDOP 493
DB 470 TWELPDGMRK 479

RESULT 5

US-09-739-254-69
Sequence 69, Application US/09739254
Patent No. US20010021700A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 49 Human Secreted Proteins
FILE REFERENCE: P2032P1
CURRENT APPLICATION NUMBER: US/09/739,254
PRIOR FILING DATE: 2000-12-19
EARLIER APPLICATION NUMBER: 09/511,554
EARLIER FILING DATE: 2000-02-23
EARLIER APPLICATION NUMBER: PCT/US99/19330
EARLIER FILING DATE: 1999-08-24
EARLIER APPLICATION NUMBER: 60/097,917
EARLIER FILING DATE: 1998-08-25
EARLIER APPLICATION NUMBER: 60/098,634
NUMBER OF SEQ ID NOS: 170
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 69
LENGTH: 502
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (502)
OTHER INFORMATION: Xaa equals stop translation
US-09-739-254-69

Query Match
Best Local Similarity 10.6%; Score 285.5; DB 10; Length 502;
Matches 124; Conservative 86; Mismatches 210; Indels 109; Gaps 22;

QY 15 LPSLLSLFLIL-----LKRNRKRTNLPKSGMPFLEGTGYLKPYATT 64
DB 14 LGGALFLILFLGLVROLKOR--PMGFPPGPKGLPIGNI-----YSLAASSELPHV 64
QY 67 PMOQSVXGKIYNSNLFGEPTIVSADAGNRFILONEGRLEFESY-----RSTIG 118
DB 65 YMRQSOVYGEISLIDGGISYVAVGVYKELVHSEIF-ADRPCLPLFMKWTMG 123
QY 119 ILGK-----NSMLVLVDMDRMRKRSISLNFISSHARLRTILKDYERHTLFVLDGMOONSI 173
DB 124 LNSRYGRGV-----DHRRLAVNSFRYFGYGO-----KSPESKIL----- 164
QY 174 FSAODEKK--PTFMKAKHMS-----MDPE-----EETE 203
DB 165 FNAITTYIGRPDRKQITNAVSNITMLIFGERFYEDIDQHMTELESENVELAASA 224
QY 217 VSAPLN-----LPGTAVHAKLOSRAITLKERKMEERKIDIREDDDEEEVKTEDA 269
DB 225 SVFLNAPFWIGILIFPGHQDLFRNAAVYDFLSRLTEKASVNRKPOLQHPVDAYLDEM 284
QY 270 ESKSDHVRKORTDDLLGWLKHSNLSSTBOILDLILSLFAGHERSSVAIALAIFLQ 329
DB 285 DQGNKD-----PSTFSKENTLIFSGELINGTETITNVLMWALIFMAL 328
QY 330 CPKAVEELREHLEIARAKKELGSEELNMDYKMDPQCVINETLRIGNVVRF-LHRKA 386
DB 329 YPNIQGOVORE-IDLIMGF-----NGKPSMDKCKMPEYTAIVAEHLRRCNIVPLGIFHAT 383
QY 389 LKDVRYGYDIPSGWKVLPVISAHLNDRYDOPNLFNPMWMOONNGASSGSGSFSTW 448

Db 384 SEDAVVGVGSIPIKGTIVITNLVSVHFEDEKVMRDEVFHBERF-----LDSGGYFAK 434
QY 449 GNNVMPFGGPRLCAGSELAKLEMAVFIHHLV-----LKFNMELAEDDOP 493
Db 435 KEALVPEFSLGRRCGELGHEHARMEMFLFTALLORFHLHPHLEVPDLKP 483

RESULT 6

US-09-904-615-69
; Sequence 69, Application US/09904615
; Patent No. US20020026040A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/904,615
; PRIOR FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 69
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (502)
; OTHER INFORMATION: xaa equals stop translation
US-09-904-615-69

Query Match 10.6%; Score 285.5; DB 10; Length 502;
Best Local Similarity 23.4%; Pred. No. 2.6e-15;
Matches 124; Conservative 86; Mismatches 210; Indels 109; Gaps 22;

QY 15 LPSLSLFLFLI---LKRNRKTRFNLPGKSGMPVLEGTIGYKPYTATTLGD---- 66
Db 14 LQGAFLFLFLFALGVRLQKRR---PMGPPGPPGLPIFGNI-----YSLAASPLPHY 64
QY 67 FMOQHVKYKGIYRSLFGEPTIVSADAGLNRFILQNEBRLFECSYR-----RSTGG 118
Db 65 YMRKSOVYGEIYFSLDGIISTVVLNGYDVKECVHSEIR-ADRPCCLPLFMKMTKMG 123
QY 119 ILGK-----WSMLVYGVGDHMRDRSISLNFSLHARLRTILKDVERTLFLVDSMOQNSI 173
Db 124 LLSNRVGRWV-----DHRRLAVNSFRYPGCG-----KSFESKIL-----EETKF 164
QY 174 FSAQDEAKK---FTFNLAKHIMS-----MDPGE---EETE---QLKKEYVFMKGV 216
Db 165 FMDALETYKGRPFEDFKQLITTNVAVSNITNLIIFGERFETEDTOFOHMIELFSENVELASA 224
QY 217 VSAPLN-----LPGTAHKALOSATILKFIKMERKADIKEDOEDEEVEYKTEDA 269
Db 225 SVFLYNAFPGIILPPGKHQOLFRNAAYVDLFLRLEKASVNRKQPLQHFVDAVIDEM 284
QY 270 EKSQSDHVKORTDDLLGLWVLKHSNLSTEQIIDLILSLFGHETSSVAIALAIFELQA 329
Db 285 DQGNKD-----PSTFSKENLIFSVGELLINGETTTTAVLMAILFMA 328
QY 330 CPKAVEELREHLEIARAKKELGESELNMDYKKMDFTQCIVINETLRIGNVVR-LHRRK 388
Db 329 YNIGGOVQKE-IDLIMR---NGKPSWDDCKKMPYTEAVLHEVLRFCNIVPLGIFHAT 383
QY 389 LKDVYKKGIDISGKVLPLVSAVHLDSKRYDQPLNFNRMQOONNGASSGSSGSSFTW 448
Db 384 SEDAVVGVGSIPIKGTIVITNLVSVHFEDEKVMRDEVFHBERF-----LDSGGYFAK 434
QY 449 GNNVMPFGGPRLCAGSELAKLEMAVFIHHLV-----LKFNMELAEDDOP 493

Db 435 KEALVPEFSLGRRCGELGHEHARMEMFLFTALLORFHLHPHLEVPDLKP 483

RESULT 7

US-10-067-668-10
; Publication No. US20030022334A1
; Sequence 10, Application US/10067668
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria Alexandra
; TITLE OF INVENTION: 33312, 33303, 32579, NOVEL HUMAN
; TITLE OF INVENTION: CYTOCHROME P450 FAMILY MEMBERS AND USES THEREOF
; FILE REFERENCE: 10448-136001
; CURRENT APPLICATION NUMBER: US/10/067,668
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/266,140
; PRIOR FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 10
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-10-067-668-10

Query Match 10.3%; Score 275; DB 9; Length 496;
Best Local Similarity 23.6%; Pred. No. 1.8e-14;
Matches 126; Conservative 82; Mismatches 208; Indels 118; Gaps 17;

QY 40 PEGKSGMPFLGELIGYK-----PYTATLGDPMQOHVSKYK-IRSLNFGEPITVSAD 93
Db 1 PEGPPLPLIGNLQGRAPGPIPHSLTKL-----RKARYGKPVPTLVIGRPVVVLG 55
QY 94 AGLNRFILQNEGRL-----FECSYRSISGILGKSMVLVYGVGDHMRDRSISLNF-- 144
Db 56 PEAVKRVLLDKGEFPAKGDNRFTPTPLSKYREGGLFSDNGPRKRLRRESLTLTF 115
QY 145 -----SHARLRTILKDVERTLFLVDSMOQNS-IFSAODEAKKFTFNLAKHIMS 194
Db 116 HFGMAYSKRSQKLEPRIOEBARDVLRIRKQAGSPIDITELLARLAPLVNVCISLFG 175
QY 195 MDGGEETQLKKEYTFPMKGVSAPLN-----LPGTAHKA 232
Db 176 V-----RFYLRPEDEFLK-LIDKLNLMPDRVSPWQOLDIFPRLYVLPGSLFRKAF 229
QY 233 GSRATILKFIKMER-----RKLDIKEDOEDEEVEYKTEDAEMKSDHYRQR 281
Db 230 KAAKLDKLDLIERRETLEPADPRIDIGFLDLSLEAKRKGANKSE----- 281
QY 282 TDDDLGLWVLKHSNLSTEQIIDLILSLFAGHETSSVAIALAIFLOACPKAVEELREH 341
Db 282 -----LSDEELATVLDLIFAGHETSSVLSMALYLLAKHREVOAKLREI 327
QY 342 LEIARAKKELGESELNMDYKKMDFTQCIVINETLRIGNV-RELIRKALKDVR-KGIYI 399
Db 328 DEVIGDR---SPYDVDAQRAQMPYIDAVIKETLRLYPVPLLPVAVAKDEIPGUYI 384
QY 400 PEGKVLPLVSAVHLDSKRYDQPLNFNRMQOONNGASSGSSGSSFTW 459
Db 385 PKGTLVIVLVYLLHDPKTFPMPPEEDPERFLDENKFKKSTA-----FLPGAGP 435
QY 460 RLCAGSELAKLEMAVFIHHLVLFKFMWELAEDDOPAFPEVDEPN-----GLP 506
Db 436 RNCIGERLARMBELFLATLQRF-----PLELAVPPGDIPSLTPPKDELGP 483

RESULT 8

US-09-945-301-4
; Sequence 4, Application US/09945301
; Patent No. US20020090699A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.


```

; APPLICANT: Glucksmann, Maria
; APPLICANT: Tsai, Fong-Ying
; TITLE OF INVENTION: 27439, NOVEL HUMAN HYDROXYLASE AND USES
; FILE REFERENCE: 38155-20036.00
; CURRENT APPLICATION NUMBER: US/09/945,301
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/229,301
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid sequence
US-09-945-301-4

```

```

Query Match 10.3%; Score 275; DB 10; Length 496;
Best Local Similarity 23.6%; Pred. No. 1,8e-14;
Matches 126; Conservative 82; Mismatches 208; Indels 118; Gaps 17;

```

```

QY 40 PRGKSWPPLGEGTIGYK-----PYATITLGDPMQOHVSKYK-ITRSNLFGEPTIYSAD 93
DB 1 PRGPPPLPLIGNLQGRAPGPIPHSLTKL-----RKARKYKRPVTLXIGPPVYVLTG 55
QY 94 AGINRFLINEGR-----FECSTPRSIGILGKMSMLVYGDHMRDRSISLNF-- 144
DB 56 PEAVKEVLIDKGEFPAKGDENFTFPMWSKGRREGILFSDGPPKRIKRSLSLTIRF 115
QY 145 -----SHARLTILKDYERHTLVLYDSMOONS-IFSADDEAKKFTFNLMARHMS 194
DB 116 HFGMGAYSRKSOKEPRIOEFARDLYERLRKQAGSPIDITELIARLAPLVANICSLIFG 175
QY 195 MDGGEETQKKEYYTEPMKGVYSAPLN-----LPGTAYHKAL 232
DB 176 V-----RFDLRPEDEPEFLK-LIDKILNEMEDRVSPMHOLLDFPPLLRYLPGLSRKAR 229
QY 233 OSRATILKFERKME-----RKIDKEEDQEEVEVTEDEAKMSKSDHYKQR 281
DB 230 KAKAKLDKVIDKIEERRETLERAGDPRRLDIFGLDSLLLEAKREGNPKSE----- 281
QY 282 TDDDLGLWVKHSNSTEQIILILSLFAGHETSSVAIALAIFLOACPRAVEELREH 341
DB 282 -----LSDELAATVLDLIFAGTETSTLSMALYLLAKHPEVQAKLREEL 327
QY 342 LETARAKKEGESELMWDYKKMDFTQCVINETLRGLNRY-RELRKALKDVAK-KGYDI 399
DB 328 DEYIGDR--SPTTYVDARAQMPYDAVYKETLRYPVYVPLLPVATKDTETIPDGYL 384
QY 400 PSGWKVLPVISAVALNRSRDQPNLFNPMWMOOQNNNGASSGSGSFTWGNMTPFGGGL 459
DB 385 PKGTIVYVNLISLRDPKYPNPEEDPERFLDENGKFKKSYA-----FLPGGAGP 435
QY 460 RLCAGSELAKLEMAVFIHLVLKFNWELADDDPFAFPVDFPN-----GLP 506
DB 436 RNCIGERLARMELFLATLILQRF-----PELELAVPPDIDISLTPKPELGLP 483

```

```

RESULT 9
US-10-067-668-12
; Sequence 12, Application US/10067668
; Publication No. US20030022334A1
; GENERAL INFORMATION:
; APPLICANT: Glucksmann, Maria, Alexandra
; TITLE OF INVENTION: 33312, 33303, 33259, NOVEL HUMAN
; FILE REFERENCE: 10448-136001
; CURRENT APPLICATION NUMBER: US/10/067,668
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 60/266,140
; PRIOR FILING DATE: 2001-02-02

```

```

; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-10-067-668-12

```

```

Query Match 10.1%; Score 272; DB 9; Length 470;
Best Local Similarity 28.6%; Pred. No. 3e-14;
Matches 87; Conservative 47; Mismatches 110; Indels 60; Gaps 8;

```

```

QY 223 LPGTAYHKALOSRATILKFERKME-----RKIDKEEDQEEVEVTEDEAKM 271
DB 194 LPGLSRKARAKAKDKVIDKIEERRETLERAGDPRRLDIFGLDSLLLEAKREGNPK 253
QY 272 SKSDHYRKQRTDDDLGLWVKHSNSTEQIILILSLFAGHETSSVAIALAIFLOACP 331
DB 254 SE-----LSDELAATVLDLIFAGTETSTLSMALYLLAKHP 291
QY 332 KAVEELREHLEIARAKKEGESELMWDYKKMDFTQCVINETLRGLNRY-RELRKALK 390
DB 292 EVQAKLREEDYIGDR--SPTTYVDARAQMPYDAVYKETLRYPVYVPLLPVATK 348
QY 391 DVRT-KGYDIPSGWKVLPVISAVALNRSRDQPNLFNPMWMOOQNNNGASSGSGSFTWG 449
DB 349 DTEIPDGYILPKGTIVYVNLISLRDPKYPNPEEDPERFLDENGKFKKSYA----- 401
QY 450 NNTMPEGGPRLCAGSELAKLEMAVFIHLVLKFNWELADDDPFAFPVDFPN----- 503
DB 402 --FLPGGAGPRNCLGERLARMELFLATLILQRF-----PELELAVPPDIDISLTPKPE 453
QY 504 GLP 506
DB 454 LGLP 457

```

```

RESULT 10
US-09-992-901-2
; Sequence 2, Application US/09992901
; Patent No. US20020073446A1
; GENERAL INFORMATION:
; APPLICANT: Neff, Michael M.
; APPLICANT: Cho, Joanne
; TITLE OF INVENTION: GENETICALLY MODIFIED PLANTS HAVING
; FILE REFERENCE: SALKINS.024DV1
; CURRENT APPLICATION NUMBER: US/09/992,901
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 09/527,073
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: US 60/124570
; PRIOR FILING DATE: 1999-03-16
; PRIOR APPLICATION NUMBER: US 60/170,931
; PRIOR FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: US 60/172,832
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-992-901-2

```

```

Query Match 9.6%; Score 256.5; DB 10; Length 520;
Best Local Similarity 23.8%; Pred. No. 6.3e-13;
Matches 136; Conservative 95; Mismatches 225; Indels 115; Gaps 25;
QY 1 MFEYHHTLPLILPSLSLFLILKLR-----RNRKTRRL-PPGSGMP--FLG 50

```

```

Db 1 MEESSWPIRVL--VLSVLSLYVKGSLMRRPKIEHNSKQIRGPIHFFIG 57
Oy 51 ---ETIGVL-----KPYATATLGFMOOHVSKYKIRVSNL--FGPPT--YVSADAG 95
Db 58 NKEVLVGMKAKASHMPSPSHNLLPRVLSFY--HHMKRYGAFVFWFG--PTFRLTVADDD 115
Oy 96 LMRFLIONGRLEFECSPYPSIGILGKMSMLVVGDM---HDMRSISLNFSLHARLRI 152
Db 116 LIRETF-SKSEFEYKNEAHPVKOLEGDLISLKGEMAHHRKILISPTHEMELKLLVY 174
Oy 153 LKDYERHRLFVLSWQ-----ONSIFSAODEAKKFTFNL 187
Db 175 YKASTD---AVDKRSKLSNGEYEVYVYEMFQITLEDVYSKRTAFSGSYEDGRVFL 230
Oy 188 MAKHIMSDPGEETEOLKREYVFMKVVSAPL--NLPGTAYHKAQSRATILKEIERK 245
Db 231 QAQOMILC-----AEAFQK---VFIPGYRFPPTGRLNLSRKLDKERI--RKSILKIEER 279
Oy 246 MEERLIDKEDEEVEEYVTEDEAEKSDHYRKQRTDDLLGWYLKHSNLSTEQILDLI 305
Db 280 -ROMAIDGEGECKEPAAK-----DLGLMIOAKNVYODIVEEC 318
Oy 306 LSLFAGHETSSVALAIFLQACPKAVEELREHELEIARAKKEGSELMWDDY--KKM 364
Db 319 KSFPRAGKQOTSNLTLWTITLLSMHEMOAKARDEVLRVC-----GSDVPTKDHVYKL 372
Oy 365 DETQOVINETLGNVRFRLHRAKLDVRYKGYDIPSGMKVLPVISAHLNDSRY--DQPN 423
Db 373 KTLMSLNESTRLLPYIVATIRRAKSDVKGKIPCGTELLIPILAHHQDAIMQNDVN 432
Oy 424 LFNPRMOQNNNGASSSGSFTSGNNMPPGCGPRCLGSELAKLEMAVFIHHLVLF 483
Db 433 ENPAPRAFADGVPRAKHPVG-----FIPGLGVKTCIGONLAILQAKTLAVMIORE 484
Oy 484 NMLEADQD--PFAPFVDFPGLRIVRSRI 512
Db 485 TEHLAPTYOHAPTYMLLXPHOGCAPITFRRL 515

RESULT 11
US-09-796-138-19
; Sequence 19, Application US/09796138
; Patent No. US20020031782A1
; GENERAL INFORMATION:
; APPLICANT: Waterman, Michael R.
; APPLICANT: Bellamine, Aouatef
; APPLICANT: Podust, Larissa M.
; TITLE OF INVENTION: Mycobacterium tuberculosis cyp51 HIGH RESOLUTION STRUCTURE, POLYPEPTIDE OF INVENTION: AND
; TITLE OF INVENTION: NUCLEIC ACIDS, AND THERAPEUTIC AND SCREENING METHODS
; TITLE OF INVENTION: RELATING TO SAME
; FILE REFERENCE: Attorney Docket No. US20020031782A1 1242-17-2
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US/09/796,138
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Penicillium italicum
US-09-796-138-19

Query Match 9.4%; Score 251; DB 10; Length 515;
Best Local Similarity 22.7%; Pred. No. 1.8e-12;
Matches 120; Conservative 75; Mismatches 213; Indels 120; Gaps 21;
Oy 12 LLLPSSLLEFLILKRRNRKTRFNLPPKSGM--PFLGRTIGY-LKPYATATLGDPMQ 69
Db 21 LFLVSLVNLVNIKQILFYNRKE-----PPVVFHMLPFGISTIAVGMOPY-----QFFF 67
Oy 70 QHVSRYGKIYNSNLFGEPTIYVSADAGLNRFILQNEGRLEFECSPYPSIGILGKMSMLVY 129
Db 70 QHVSRYGKIYNSNLFGEPTIYVSADAGLNRFILQNEGRLEFECSPYPSIGILGKMSMLVY 129

```

```

Db 68 ASRAKYGDIFFELILGKKTIVYLGVEGNEFIL--NGKLVDVNAEE---VYGRITTPVFG 121
Oy 130 GDMHRMRSGISL---NPLSHARLFTILKDYERHRLFVLSMQONSIFSADQEAK----- 181
Db 122 SDVYTCPSKMLEQKFKITGLSQEAL-----ESTYPLIAD--ETNAYIKSSPNEFGQSG 175
Oy 182 -----KTFENMAKHIMSDPGEETEOLKREY-----VFPMKVVSAPL 219
Db 176 TIDLAAAMAEITFTFAARLIQ-----GEEVRSKLTSEFADLFHDLGFSPIINMLPAPL 231
Oy 220 PLN---LPGTAYHKAQSRATILKFERKMEERKLDKEDEEVEEYKTEDEAEKSDH 276
Db 232 PHNAAIKHTTYARDLSGNVPASATGSMRRQRROD----- 267
Oy 277 VRKQRPD--DDLGLVYLKHSN--LSTEQILDLISLFPAGHETSSVALAIFLQACPKA 333
Db 268 -KSKGTDMSNLKRCYRQGTPIPKKEIAHMTTLMAQOHSASLSCILLRLASQPEK 326
Oy 334 VEBLEHELEIARAKKEIGES--ELNWDYKKMDFQCVINETLRLGNVRFRLHRAKLD 391
Db 327 AERLHAEQI-----KNLGADLPRLQYKDMDKLPLLRNYIKETLRHSSIHFLMRKVKNP 380
Oy 392 VRYKGYD--IPSGMKVLPVISAHLNDSRYDQPNLEPNRMOQ-----NNGA 437
Db 381 MPVPGTDFVVPSPSHLSSPGYTARDEHFRDPLRMDPHRWESRYVEDSSDTVDYGYGA 440
Oy 438 SSSGGSFTSGNNMPPGCGPRCLGSELAKLEMAVFIHHLVLF 485
Db 441 VSKGTS-----PYLPFGAGRHCIGEFAYLNLVYATVLRERF 482

RESULT 12
US-09-909-903-19
; Sequence 19, Application US/09909903
; Patent No. US20020052031A1
; GENERAL INFORMATION:
; APPLICANT: Waterman, Michael R.
; APPLICANT: Bellamine, Aouatef
; TITLE OF INVENTION: Mycobacterium tuberculosis cyp51 POLYPEPTIDES AND
; TITLE OF INVENTION: NUCLEIC ACIDS AND THERAPEUTIC AND SCREENING METHODS
; TITLE OF INVENTION: RELATING TO SAME
; FILE REFERENCE: Attorney Docket No. US20020052031A1 1242-17
; CURRENT FILING DATE: 2001-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Penicillium italicum
US-09-909-903-19

Query Match 9.4%; Score 251; DB 10; Length 515;
Best Local Similarity 22.7%; Pred. No. 1.8e-12;
Matches 120; Conservative 75; Mismatches 213; Indels 120; Gaps 21;
Oy 12 LLLPSSLLEFLILKRRNRKTRFNLPPKSGM--PFLGRTIGY-LKPYATATLGDPMQ 69
Db 21 LFLVSLVNLVNIKQILFYNRKE-----PPVVFHMLPFGISTIAVGMOPY-----QFFF 67
Oy 70 QHVSRYGKIYNSNLFGEPTIYVSADAGLNRFILQNEGRLEFECSPYPSIGILGKMSMLVY 129
Db 68 ASRAKYGDIFFELILGKKTIVYLGVEGNEFIL--NGKLVDVNAEE---VYGRITTPVFG 121
Oy 130 GDMHRMRSGISL---NPLSHARLFTILKDYERHRLFVLSMQONSIFSADQEAK----- 181
Db 122 SDVYTCPSKMLEQKFKITGLSQEAL-----ESTYPLIAD--ETNAYIKSSPNEFGQSG 175
Oy 182 -----KTFENMAKHIMSDPGEETEOLKREY-----VFPMKVVSAPL 219
Db 176 TIDLAAAMAEITFTFAARLIQ-----GEEVRSKLTSEFADLFHDLGFSPIINMLPAPL 231
Oy 220 PLN---LPGTAYHKAQSRATILKFERKMEERKLDKEDEEVEEYKTEDEAEKSDH 276
Db 232 PHNAAIKHTTYARDLSGNVPASATGSMRRQRROD----- 267
Oy 277 VRKQRPD--DDLGLVYLKHSN--LSTEQILDLISLFPAGHETSSVALAIFLQACPKA 333
Db 268 -KSKGTDMSNLKRCYRQGTPIPKKEIAHMTTLMAQOHSASLSCILLRLASQPEK 326
Oy 334 VEBLEHELEIARAKKEIGES--ELNWDYKKMDFQCVINETLRLGNVRFRLHRAKLD 391
Db 327 AERLHAEQI-----KNLGADLPRLQYKDMDKLPLLRNYIKETLRHSSIHFLMRKVKNP 380
Oy 392 VRYKGYD--IPSGMKVLPVISAHLNDSRYDQPNLEPNRMOQ-----NNGA 437
Db 381 MPVPGTDFVVPSPSHLSSPGYTARDEHFRDPLRMDPHRWESRYVEDSSDTVDYGYGA 440
Oy 438 SSSGGSFTSGNNMPPGCGPRCLGSELAKLEMAVFIHHLVLF 485
Db 441 VSKGTS-----PYLPFGAGRHCIGEFAYLNLVYATVLRERF 482

```

OY 220 PLN---LPGTAHKAQSAFTIKETERMEERKLDIKEEDOEFEVTEDEAKMSDH 276
 DB 232 PHNASHAKHTFYARDLSGVNPSATGSMRRORROD----- 267
 OY 277 VKORRD--DILLGWLKSN--LSTEQILDILISLIFAGHETSVAIALAIFLOCPKA 333
 DB 268 -SKSGDMSLMLRCVYRGTPIPDKEIAHMMITLLMAQOHSASISWILLASOPKA 326
 OY 334 VEREHELEIARAKKELES--ELNMDYKKMDFTQCVINETRLGNVVRFLRRKALD 391
 DB 327 AEKLAHBOI-----KNIGADLEPIQYKMDKPLRLANVIKETRLRHSSHTLMRKVKNP 380
 OY 392 VRYKGD--IPSGKVLPIVISAHLDNSRYDQPNLEFPMRMOOQ-----NGA 437
 DB 381 MPYGDFTFVPPSHLTLSPPGVYARDEHRFDPRLRMDPHRESRVEVEDSDTVGYGA 440
 OY 438 SSGSGSFTWGNNTYPPGCGPRLCAGSELAKEAMVFIHHLVLRKNW 485
 DB 441 VSKGTRS-----PYLPFGAGRHCIGERFAVAILNLEVIATLVREFR 482

RESULT 13
 US-10-067-668-8
 ; Sequence 8, Application US/10067668
 ; Publication No. US20030022334A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Glucksmann, Maria Alexandra
 ; TITLE OF INVENTION: 33312, 33309, 32579, NOVEL HUMAN
 ; FILE REFERENCE: 10448-136001
 ; CURRENT APPLICATION NUMBER: US/10/067, 668
 ; CURRENT FILING DATE: 2002-02-04
 ; PRIOR APPLICATION NUMBER: 60/266, 140
 ; PRIOR FILING DATE: 2001-02-02
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 544
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 US-10-067-668-8

Query Match 9.4%; Score 251; DB 9; Length 544;
 Best Local Similarity 23.2%; Pred. No. 1.9e-12;
 Matches 127; Conservative 98; Mismatches 234; Indels 88; Gaps 23;
 OY 13 LILPLSLILFLILKRRNRKTRFNLPEKSGMPFLGRTIGY--LKPY----- 59
 DB 37 LILGLVALLGWSMLRRRRAR---GIPGPTWPLVG--NFGVLLPPLRLRRSSWLSRT 91
 OY 60 ----TATTLG-DPMOQHVSK-YGKIYRSLFGEPTIVSADAGLNRFILQNEGRLEPCSY 112
 DB 92 RANGIDPSVIGPVLALHARVGSIFSPFIGHVYVLSDFHSVRALVQAQAEVPS-DR 150
 OY 113 PR-SIGGILGKMSLVYV--GDMHRDRSISLNFSLHARLTLL--KDVERTHLEVYDS 167
 DB 151 PRVPLISIVYKKEGVFAHGPVWRQKRFSHSLRHFGLKSLLEKPIIEEK-YKAE 209
 OY 168 MOONSIFSADAEKKEFF--NIMAKHIMSMDPE--EETQOLKEVYTFKGVASAPLN- 222
 DB 210 MOKH---GDDPCPFSIISNAVSNIISLCLFQORFYTNSERKMKMGFSGRLGELTNS 265
 OY 223 -----LPGTAHKAQSAFTIKETERMEERKLDIKEEDOEFEVTEDEAKMSDH 269
 DB 266 OVLVNICPMLVYLPFGRFELROIEKDITSL-----KTIKQOESLDRENPDQFI 318
 OY 270 ESKSDHVRKORTDDLLGWVLRKSNLSTEQILDILISLIFAGHETSVAIALAIFLOQA 329
 DB 319 DMYLL-HMEERKNN-----SNSSFEDEYLFYIIGDLEIAGDTTNSILMCLVMSL 370
 OY 330 CPRAVELEIREHELEIARAKKELESSELMMDYKKMDFTQCVINETRLGNVVR-LLRKA 388
 DB 371 NPVOQEVHEIEIRVIGANR-----APSLTDKQAMPYTEATIMEVORLTJVAVPLAIPHT 425

OY 389 LADVRYKYDIPSGKVLPIVISAHLDNSRYDQPNLEFPMRMOOQNGASSSGSFSFTW 448
 DB 426 SENTVLOGYTRIGLTLILNLSVHRDPALWEKPEDFYFNRLDDQGLK----- 477
 OY 449 GNNYMPFGGPRLCAGSELAKEAMVFIHHLVLRKWELEAD-QPF---APPYDFPN 504
 DB 478 -ETTFPFGIGKRYCQGEOLAKMELFIMFVLSQSFALPDESKRPLTGRGTLAPHP 536
 OY 505 LPIVRS 511
 DB 537 ENITISR 543

RESULT 14
 US-09-796-138-18
 ; Sequence 18, Application US/09796138
 ; Patent No. US20020031782A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Waterman, Michael R.
 ; APPLICANT: Podust, Larissa M.
 ; TITLE OF INVENTION: Mycobacterium tuberculosis CYP51 HIGH RESOLUTION STRUCTURE, POL
 ; TITLE OF INVENTION: AND
 ; TITLE OF INVENTION: NUCLEIC ACIDS, AND THERAPEUTIC AND SCREENING METHODS
 ; FILE REFERENCE: Attorney Docket No. US20020031782A1 1242-17-2
 ; CURRENT APPLICATION NUMBER: US/09/796,138
 ; PRIOR APPLICATION NUMBER: 09/345,218
 ; PRIOR FILING DATE: 1999-06-30
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 18
 ; LENGTH: 503
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 US-09-796-138-18

Query Match 9.3%; Score 249; DB 10; Length 503;
 Best Local Similarity 24.1%; Pred. No. 2.5e-12;
 Matches 132; Conservative 87; Mismatches 212; Indels 116; Gaps 28;
 OY 9 LILPLSLILFLILKRRNRKTRFNLPEKSGMPFLGRTIGY--LKPY----- 60
 DB 26 LLSMLIACFTSLVLYLRLAAGH--LYQLPAGVSPYIRSPIDFGLHALAFKSP 82
 OY 61 ATTLDPMOQHVSKYKTIYRSLFGE--PTIVSADAGLNRFILQNEGRLEPCSYPSIG 118
 DB 83 ----EFLNAYERYKGVFSTYVKGKTYLIGSDAALFNSKNEDLNAMEDYYSRLTTP 137
 OY 119 ILGKMSLVYVGM---HRDMRSISLN--FLSHARLTLLKDVERTHLEVYDSMOON- 171
 DB 138 VEGKGVAYDVNPVYLFQKMKLSGLNIAHFKQVSI-----IEKETEYESMGESG 190
 OY 172 -SIFSADAEKKEFTFNLMAKHIMSMDPGEETEOLKE---YTFMKGVASAPLNLP 225
 DB 191 EKNVEALSE---LILTFASHCLH--GKEIRSQLNEKVAOLYADDGFSHAMVLLP 243
 OY 226 ----TAVHKAQSAFTIKETERMEERKLDIKEEDOEFEVTEDEAKMSDHVAKOR 281
 DB 244 WLPSPFRDRRAHREIKDIFYRAIOKR-----OSOKI----- 278
 OY 282 TDOLLGWVLRKSN--LSTEQILDILISLIFAGHETSVAIALAIFLOCPRAVEE 336
 DB 279 --DDILOTLDATYKORPLTDEVAGMLGILLAGHTSSTISAMWGFL-ARDKTLQ- 334
 OY 337 LREHELEIARAKKELES--ELNMDYKKMDFTQCVINETRLGNVVRFLRRKALDVR 394
 DB 335 -KCYLE--OKVCGENLPLVYDQKDLNDRICKEFRLRBPPIIMWMAKRPQTV 390
 OY 395 KGDIPIGKVL--PVSVAHLDN--SRIDQPNLEFPMRMOOQNGASSSGSFSFTW 450

Db 391 AGTTPRGHVCSPVNO RLKDSWVERLD-----FNPDRILODN---PASGE-KFA----- 438
QY 451 NMPFGGPRRLCAGSELAKLEMAVFIHLVLKFNWELAEODPFAFPFVDF-----PNG 504
Db 439 -YVPGAGRHRCIGENFAVVAQIKTIWSTMLRIYERDLIDG-----YFPVNYTMTIHTPEN 493
QY 505 LPIRYSR 511
Db 494 PVIRYKR 500

Db 494 PVIRYKR 500

Search completed: March 30, 2003, 12:14:58
Job time : 54 secs

RESULT 15
US-09-909-903-18
; Sequence 18, Application US/09909903
; Patent No. US20020052031A1
; GENERAL INFORMATION:
; APPLICANT: Waterman, Michael R.
; TITLE OF INVENTION: Mycobacterium tuberculosis CYP51 POLYPEPTIDES AND
; TITLE OF INVENTION: NUCLEIC ACIDS AND THERAPEUTIC AND SCREENING METHODS
; FILE REFERENCE: ATTORNEY DOCKET NO. US20020052031A1 1242-17
; CURRENT APPLICATION NUMBER: US/09/909,903
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-903-18

Query Match 9.3%, Score 249; DB 10; Length 503;
Best Local Similarity 24.1%, Pred. No. 2.5e-12;
Matches 132; Conservative 87; Mismatches 212; Indels 116; Gaps 28;

QY 9 LIPILLPSLST-LFLILLKRRNRKTRFNLPCK-----SGMPFLGRTIGYLRPT 60
Db 26 LLSMLLCAFLSLVYLRLAAGH--LYOLPAGVKSPYIESPIPLGHAIAFGKSPI 82
QY 61 ATTLDGFMQOHVSKYKGYRSNLGCE--PTIYSADAGLNRFILQNEGRLEFCSYPRSIG 118
Db 83 -----EFLNAYEKYGFPSFTMGKFTYLLGSDAALLFNSKMDLNAEDVYSLTTP 137
QY 119 ILGKSMVLVYGD--HRDMSISLN--FLSHARLRTILKDYERHRTLEVLDWMQN- 171
Db 138 VEGKVAADVDPVPLDQKMKLSGLNIAHFQHYSI-----LEKETKEFEESMGESG 190
QY 172 --SIFSADDEAKKFTFNLMAKHIMSDGDEETEDLKE---YTFMKGVYSAPLNLPG 225
Db 191 EKNVEALSE---LIITLASHCLH--GKEIRSQLNEKVAQLVADLDGGSFHAAMLPG 243
QY 226 ---TAYHKALOSRAATILKFERKMEERKLDIKEDEEEVKTDEAEKMSKSDHVRKOR 281
Db 244 WLPSPSFRRRDRRAHREIDIFYKAQKRR-----QSOEKT----- 278
QY 282 TDDLLGVNLKHS-----NLSTEQILDLISLFGHETSSVALALAFLOACPKAVEE 336
Db 279 --DDILQTLILATYKDGRLPLTDEEAGMLIGILLAGQHTSSTTSAMGFLL-ARDKTLQ- 334
QY 337 LREHLEIARAKKELGES--ELNMDYKKMDTQCIVINETLRIGNVAVFELHRAKALKDVRX 394
Db 335 -KKCYL--OKTVCGENLPLTYDOLKDLNDRICKETRLRPPIMIMRMARTPQTV 390
QY 395 KGYDIPSGMKYL--PVISAVHLDN--SRYDOPNLFPNPMOQOONNGASSGSGSFTWGN 450
Db 391 AGTTPRGHVCSPVNO RLKDSWVERLD-----FNPDRILODN---PASGE-KFA----- 438
QY 451 NMPFGGPRRLCAGSELAKLEMAVFIHLVLKFNWELAEODPFAFPFVDF-----PNG 504
Db 439 -YVPGAGRHRCIGENFAVVAQIKTIWSTMLRIYERDLIDG-----YFPVNYTMTIHTPEN 493
QY 505 LPIRYSR 511

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:

March 30, 2003, 12:04:26 ; Search time 49 seconds
(without alignments)
1006.469 Million cell updates/sec

Title:
US-09-502-426a-2

Sequence: 1 METEHTLPLPLPLPSLS.....FAFPVDPNGLPIKVSRL 513

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

PIR_73:*
1: pirl1:*
2: pirl2:*
3: pirl3:*
4: pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length DB	ID	Description
1	2677	99.9	513	2	T46143
2	978.5	36.5	472	1	S55379
3	866	32.3	512	2	H96759
4	860	32.1	457	2	D85429
5	815	30.4	382	2	T48613
6	761.5	28.4	464	2	T07859
7	632	23.6	457	2	T04444
8	611	22.8	482	2	T02739
9	581	21.7	487	1	C71417
10	579.5	21.6	485	2	A84859
11	579.5	21.6	485	2	H86185
12	577	21.5	455	2	T48973
13	547.5	20.4	489	2	B84733
14	497.5	18.6	460	2	D96813
15	491.5	18.3	519	1	T02263
16	489.5	17.9	444	1	T04602
17	479.5	18.3	369	1	S75761
18	369	13.8	349	2	A86329
19	357	13.3	518	2	T20908
20	338	12.6	453	2	C83722
21	329.5	12.3	517	2	T20907
22	324.5	12.1	520	2	T24778
23	321.5	12.0	518	2	T24783
24	316	11.8	500	2	T24737
25	310.5	11.6	520	2	T24777
26	306.5	11.4	500	2	T52175
27	302.5	11.3	504	2	A25222
28	300	11.2	491	2	S31277
29	300	11.2	491	2	I84735

30	299.5	11.2	492	2	S27160	cytochrome P450 2B
31	297	11.1	491	2	A31047	testosterone 16alp
32	297	11.1	494	2	A33293	cytochrome P450 2A
33	294.5	11.0	491	1	O4RBP3	cytochrome P450 2B
34	294	11.0	506	2	D96672	probable Cytochrom
35	294	11.0	516	2	T48140	flavonoid 3',5'-hy
36	292	10.9	491	1	O4RBP3	cytochrome P450 2B
37	291.5	10.9	491	2	S35666	cytochrome P450 2B
38	291.5	10.9	530	1	A27491	lanosterol 14alp
39	290.5	10.8	502	2	T05246	cytochrome P450 9
40	288.5	10.8	504	2	A29410	cytochrome P450 1
41	288	10.7	459	2	A61967	cytochrome P450 3A
42	287.5	10.7	520	2	T24780	hypothetical prote
43	287	10.7	503	2	JC4702	cytochrome P450 3A
44	284.5	10.6	505	1	S38534	cytochrome P450 76
45	284	10.6	501	2	T04735	cytochrome P450 76

ALIGNMENTS

RESULT 1

T46143

steroid 22-alpha-hydroxylase (DMF4) - Arabidopsis thaliana

N:Alternate names: protein T3A5.40

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 16-Feb-2001

C:Accession: T46143

R:Biocheck: H.; Mewes, H.W.; Mayer, K.F.X.; Lemcke, K.; Schaefer, C.; Quettler, F.; S

submitted to the Protein Sequence Database, December 1999

A:Reference number: 223024

A:Accession: T46143

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-513 <BLO>

A:Cross-references: EMBL:AL132979

A:Experimental source: cultivar Columbia; BAC clone T3A5

A:Map position: 3

A:Insertions: 74/2; 182/3; 233/3; 338/3; 369/3; 396/1; 432/3

A:Note: T3A5.40

C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: heme; iron; metalloprotein

F:308-484/Domain: cytochrome P450 homology <P45>

F:462/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 99.9%; Score 2677; DB 2; Length 513;
Best Local Similarity 99.8%; Pred. No. 1.1e-165;
Matches 512; Conservative 1; Mismatches 0; Gaps 0;

OY	1	METEHTLPLPLPLPSLSLTLFLILKRRNRKTRFNLPGKSGWPLGRTIGLYKPYT	60
DB	1	METEHTLPLPLPLPSLSLTLFLILKRRNRKTRFNLPGKSGWPLGRTIGLYKPYT	60
OY	61	ATTLDPMQOHSKYGKTYRSLFGEPTIVSADAGLNFTILONEGRLECYSPISGIL	120
DB	61	ATTLDPMQOHSKYGKTYRSLFGEPTIVSADAGLNFTILONEGRLECYSPISGIL	120
OY	121	GKMSLVLVGDHNRDMRSISLNFSLHARLRTILKDVRRHLFLVDSWQONSIFSADDEA	180
DB	121	GKMSLVLVGDHNRDMRSISLNFSLHARLRTILKDVRRHLFLVDSWQONSIFSADDEA	180
OY	181	KFTFNMAKTHMSDPEEETEDLKKETVTFMGVVSAPLNLPGTAVHKAJOSRATILK	240
DB	181	KFTFNMAKTHMSDPEEETEDLKKETVTFMGVVSAPLNLPGTAVHKAJOSRATILK	240
OY	241	FLERMEERKLDIPEEDDEEVKTEDEAEMSKSDHVRKQRTDDLLGWYKHSNLSSTEQ	300
DB	241	FLERMEERKLDIPEEDDEEVKTEDEAEMSKSDHVRKQRTDDLLGWYKHSNLSSTEQ	300
OY	301	ILDLILSLFAGHETSSVAIALAIFFLQACPKAVEELREELFLIARAKKELGSEELNWD	360
DB	301	ILDLILSLFAGHETSSVAIALAIFFLQACPKAVEELREELFLIARAKKELGSEELNWD	360

OY 361 YKKMPTQCVITETRLGKVVAFELHKKALKDVRYKGYDIPSGMKVLPVISAHLNDRYD 420
DB 361 YKKMPTQCVITETRLGKVVAFELHKKALKDVRYKGYDIPSGMKVLPVISAHLNDRYD 420
OY 421 QPNLNPMMQOONNGASSGSGSFSTWGNMMPFGGGRCLCAGSELKLEMAVFIHLHY 480
DB 421 QPNLNPMMQOONNGASSGSGSFSTWGNMMPFGGGRCLCAGSELKLEMAVFIHLHY 480
OY 481 LKFMNELAEDDQFAFPFVDFPNGLPPIRYSRL 513
DB 481 LKFMNELAEDDQFAFPFVDFPNGLPPIRYSRL 513

RESULT 2

S55379
cytochrome P450 CYP90 - Arabidopsis thaliana
N.Contains: oxidoreductase (EC 1.-.-.-)
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C.Accession: S55379
R.Szekeres, M.; Nemeth, K.; Koncz, Z.; Nagy, F.; Koncz, C.
submitted to the EMBL Data Library, May 1995
A.Reference number: S55379
A.Accession: S55379
A.Molecule type: mRNA
A.Residues: 1-472 <SZR>
A.Cross-references: EMBL:X87367; NID:9853718; PIDN:CA60793.1; PID:9853719
C.Genetics:
A.Gene: CYP90
C.Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C.Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;
F:275-440/Domain: cytochrome P450 homology <P45>
F:418/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 36.5%; Score 978.5; DB 1; Length 472;
Best Local Similarity 41.9%; Pred. No. 7.5e-56;
Matches 210; Conservative 79; Mismatches 171; Indels 41; Gaps 8;

OY 12 LLLPSLSLLFLILKRRNRKTRFNLPPKSGWPFLEGTIGYLPYATTLGDFMOOH 71
DB 7 LLLSSIAAGFL--LLLRTRRYRMGLPPGSLGLPGLGETFQLGAVKTEPPEFIDER 63
OY 72 VSKYKIRTSNLFGEPTIVSADAGLNRLPLONEGRLEFCSYRISGIGLKRSMLVYGD 131
DB 64 VARYSSVMTLHFGEPITFSADPELNRFVLONEGRLEFCSYRISGIGLKRSMLVYGD 123
OY 132 MHRDRSISLNFILSHARLRTILKDVRRHTLFLVDSMOONSIFSAODEAKKRTFMMAKH 191
DB 124 LHKRHSILTMSPANSITIKDMLMDIDRLVRFNLDSMSRYLL--MEBAKITEFELVQ 181
OY 192 IMSDNGEETEQKKEYVTPEKGVASAPLPGTAHYKALQSRAITILFERKKEERKL 251
DB 182 LMSFDPG--EWSRSLRKEVLIVTEGFSPLPLPFSITRYKALQAR-----RKVAELT 232
OY 252 DIKEDQEEBEKVTDEDEMSKSDHVRKQRTDDLLGWLKHSNISTQIIDLIIISLPA 311
DB 233 VYVAKRREEEBGAE-----RKDKMLAALIAADGFSDEIDVFLVALLVA 278
OY 312 GHETSSVAIALAIFLOACPAVEELREHLEIARAKKEGSELMNDYKMDTQCVI 371
DB 279 GYETISITMTLAVKLTETPLALADLKEHEKIRAKMSD--SYLSMSDYISMPDQCV 336
OY 372 NETLRLGNVVFELHKKALKDVRYKGYDIPSGMKVLPVISAHLNDRYDQPLPMPAMQ 431
DB 337 NETLRLGNVVFELHKKALKDVRYKGYDIPSGMKVLPVISAHLNDRYDQPLPMPAMQ 396
OY 432 QONNGASSGSGSFSTWGNMMPFGGGRCLCAGSELKLEMAVFIHLHYLKNELAED 490
DB 397 -----SNSTYTGFSNVETFPFGGGRCLCAGSELKLEMAVFIHLHYLKNELAED 446
OY 491 DQFAFPFVDFPNGLPPIRYSRL 511

DB 447 DKLVFEPPTTRQKRYPIFYKR 467

RESULT 3

H96759

probable steroid 22-alpha-hydroxylase T9L24.44 [imported] - Arabidopsis thaliana
C.Species: Arabidopsis thaliana (mouse-ear cress)
C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Apr-2002
C.Accession: H96759
R.Theologis, A.; Ecker, J.R.; Palm, C.J.; Federle, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A.Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A.Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A.Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A.Reference number: A6141; MUID:21016719; PMID:11130712
A.Accession: H96759
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-512 <STO>
A.Cross-references: GB:A605173; NID:q11120803; PIDN:AA630983.1; GSPDB:GN00141
C.Genetics:
A.Gene: T9L24.44
A.Map position: 1
C.Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C.Keywords: heme; iron; metalloprotein
F:445/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 32.3%; Score 866; DB 2; Length 512;
Best Local Similarity 33.8%; Pred. No. 1.6e-48;
Matches 181; Conservative 105; Mismatches 173; Indels 76; Gaps 7;

OY 12 LLLPSLSLLFLILKRRNRKTRFNLPPKSGWPFLEGTIGYLPYATTLGDFMOOH 71
DB 11 LLSVSSSTFLAFIIFLAGIARRRRAPRLPGSGMWPLIGDTPFAMLVNAGSHSS 70
OY 67 FMOQVSK-----YKIRTSNLFGEPTIVSADAGLNRL 101
DB 71 FVEKIRKFFVSLCYLILKRPDNGSFNEIRYRITSCSLFGMAVAVSADPPENRITM 130
OY 102 ONEGRLEFCSYRISGIGLKRSMLVYGDIMHRDRSISLNFILSHARLRTILKDVRRHT 161
DB 131 ONEGRLEFCSYRISGIGLKRSMLVYGDIMHRDRSISLNFILSHARLRTILKDVRRHT 190
OY 162 LFLVDSMOONSIFSAODEAKKRTFMMAKHISMSPGEEETQKKEYVTPEKGVASAPL 221
DB 191 LQTSNFFDGEVLLQDICRKAHILMAYNOLGVS--SESEVDEMSQLPSDFVDCGLSVPI 249
OY 222 NLPGTAHYKALQSRAITILFERKKEERKLQKEDQEEBEKVTDEDEMSKSDHVRKQ 281
DB 250 DLPGETYKMAKARKETIRKINKTTEKRLONKASD-----T 286
OY 282 TDDDLGWLKHSNISTQIIDLIIISLFAHETSSVAIALAIFLOACPAVEELREHLE 341
DB 287 AGNGVILGRLEESLPNSMADFIINLFPAGNETSKMLFAVYVLTGCPAMQOLLEH 346
OY 342 LEIARAKKEGSELMNDYKMDTQCVIETLRLGNVVFELHKKALKDVRYKGYDIPS 401
DB 347 -----DRLAGMLTMDQYKMDTQCVIETLRLGNVVFELHKKALKDVRYKGYDIPS 399
OY 402 GMKVLPIVSAHLNDRYDQPLPMPAMQ-----QONNGASSGSGSFSTWGNMMPFG 456
DB 400 GCFVVPFLSAVHLNDRYKESLSFSPRMLDPEYQKRNMTSP-----FYCFPG 449
OY 457 GGPRLCAGSELKLEMAVFIHLHYLKNELAEDDQFAFPFVDFPNGLPPIRYSRL 511
DB 450 GSTRCPGAEALRQIALFHLHYFITTYKMTQKEDRISIFFFSARLYVNGFKQLNR 504

RESULT 4

D85429

cytochrome P450 like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Apr-2001

C:Accession: D85429

R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617198

A:Accession: D85429

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-457 <STO>

A:Cross-references: GB:NC_001268; NID:97270586; PIDN:CA80304.1; GSPDB:GN00140

A:Gene: AT4g36380

A:Map position: 4

C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

F:396/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match

Best Local Similarity 32.1%; Score 860; DB 2; Length 457;

Matches 179; Conservative 93; Mismatches 160; Indels 54; Gaps 8;

39 LPPGKSGMPLEGTIGYL-----KPYATTTGDPWQOHVSKYKITYNSLFGPEPTVSA 92

2 TRNGSLGMPVIGETLNFICGYSSRPVT-----FMDKRKSLGYGVFKNIIGTPIIST 55

93 DAGLNFTLQNBGRLEFESYPSRISGILGKMSLVLVGDMHDMDSILNPLSHARKTI 152

56 DAENVKVVYLVQNGNFTVPRPSTITELGENSEILSINGPHOKRITLIGAFIRSHLDOR 115

153 LKQVNRHTLVLDWQONSIFSAODEAKFTFNLMANHMSMDGEEETDLKKEYTTF 212

116 TRDIEASVYLTASMAOLPLVHODEIKMTFELIVYLMSTSPG-EDMTLKLEFEFF 174

213 MKGVASAPLNLPGTAYHKAALSRATILFKERKMERKLDIKEDEEVEEKT-----E 266

175 IKGILCIPKPRGTLYKSLAKKEKLIKMKVVEKRYAMTTSPANDVDVLLRDGD 234

267 DEAKESKSDHVRKQRTDDLLGWLKHSNLSLSTEDLLILSLFAGHETSSVAIALATFF 326

235 SEKQSPDSDVSGK-----IYEMMIPGEETMTPAMTLANKF 270

327 LOACRAVEBELREHLEIARAKKEISELNMWDYKKADFTQCYINFTLGLGNVRLTR 386

271 LSDNPAVALAKVEEMKRRKRLLEGE-EYKMTDYSLSFTQNYINETLRMANIINGVWR 329

387 KALADVRKGYDIPSGMKVLPVISAHLNDSRYDOPNLFNPRMOQONNGASSSGSFS 446

330 KALKDVEIKGYLIPKGMCVLASFISVHDEDIYNPOFDPWRDRINGSNSSIC---- 385

447 TWGNWYMPFGGPRICAGSELAKLEMAVFIHLVLFKFMELAEDDOPAPFPVDFPGLP 506

386 -----FPPFEGGGRICDGLLESLKLEISIFLHLVTRYSW-TAEDEIYSPFVAMKRRLP 439

507 IRVRSRI 512

440 IRVATV 445

RESULT 5

T48613

hypothetical protein F18022.190 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48613

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le

submitted to the Protein Sequence Database, April 2000

A:Reference number: 224493

A:Accession: T48613

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-382 <BEV>

A:Cross-references: EMBL:DB

A:Experimental source: cultivar Columbia; BAC clone F18022

C:Genetics:

A:Map position: 5

A:Introns: 31/2; 139/3; 204/1; 224/3; 253/3; 280/1; 315/3; 349/2

A:Note: F18022.190

Query Match

Best Local Similarity 30.4%; Score 815; DB 2; Length 382;

Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

45 GMPFEGTIGYKPYTATTTGDPWQOHVSKYKITYNSLFGPEPTVSDAGINREITONE 104

2 GMPFEGTISFEKPRHSISGTFLOQORSRYGVKYSNCGKAVSCDQELNMFILQNE 61

105 GRLPFCYSPRSISGILGKMSLVLVGDMHDMDSILNPLSHARKTILLADVERHTLFV 164

62 GRLFTSDPKAMHDLGKYSLLATGEIRKIKNYISFIMTKSPDLHCANLSISI 121

165 LSWQONSIFSAODEAKFTFNLMANHMSMDGEEETDLKKEYTTFMKGVASAPLNP 224

122 LKSMKNCREVEFHKKEVKITFLSVWVQNLISIKEDPARLVYLDPLSMKGFISLPDLP 181

225 GYAVHKAOSRATILFKERKMERKLDIKEDEEVEEKTDEDAESKSDHVRKQRTDD 284

182 GTGTYNAIKVSNRNITHONALIEDNNALIREDDLSISNED----- 224

285 DLGQVLRKSNLSLEQJLILSLFAGHETSSVAIALATFFLOACRAVEBELREHLEI 344

225 -----EEHNAI 230

345 ARAKKEISELNMWDYKKADFTQCYINFTLGLGNVRLTRKALDKVRRKYDIPSGWK 404

231 -RANKGDEL-LNNEEDYQKMEFTQCVISEALRCNINIKYVARKATHIDIKENEYIIPGKW 288

405 VLPYISAVHLNDSRYDOPNLFNPRMOQONNGASSSGSSTMGNNYMPFGGPRICAG 464

289 VEPFTAVHLDPSLHNEFENPRMWTKT-----AFGGGVRCVPG 329

465 SELAKLEMAVFIHLVLFKFMELAEDDOPAPFPVDFPGLP 509

330 GELGKLIAPFLHILVLYSRWKIKSDPMPIAHPIVVEKRGMLLEI 374

RESULT 6

T07859

cytochrome P450 homolog - tomato

C:Species: Lycopersicon esculentum (tomato)

C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 16-Feb-2001

C:Accession: T07859

R:Bishop, G.J.; Harrison, K.; Jones, J.D.

Plant Cell 8, 959-969, 1996

A:Title: The tomato Dwarf gene isolated by heterologous transposon tagging encodes th

A:Reference number: T16181; MUID:9626705; PMID:8672892

A:Accession: T07859

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-464 <BIS>

A:Cross-references: EMBL:U54770; NID:g1421740; PIDN:AA817070.1; PID:g1421741

A:Experimental source: strain GCR758

C:Genetics:

A:Gene: dwarf

C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

F:273-436/Domain: cytochrome P450 homology <P45>

Query Match

Best Local Similarity 28.4%; Score 761.5; DB 2; Length 464;

Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

12 LILLPILSLILLFLILKRRN--RKTFRNLPPGKSGMPLEGTIGYKPYTATTTG-DPM 68

DB 5 LFLSFEELCJFCTALHLMNOVKYKQKLPPTGMWPLFGETTEK-----LDPSPM 58
OY 69 OOHVSKYKTYNSNLFGEPTIVSADAGLNRFILIONBGRLECSYPRISGILGKMSMLV 128
DB 59 KNOBARVSGFESHILGCTIYSMOSELNRYILYNNAKGLVPGQSMIDILGKCAIYV 118
OY 129 VQDMHEDMSISLNFSLHARLTKLKYERHTLVYDSMOONSIFSADEAKKFTFNLM 188
DB 119 NOSAHRYMGALLSLSPIMIRDLPKIDEMRSHLTWM-DNKVIDIEKTKMAFLSS 177
OY 189 AHHNSMDGEEET---EOLKKEVYFMKGVYAPNLGTAVHAKLOSATILKIEK 245
DB 178 LKQI---AGISTSLAQFNSFFVLVLTSLPILNLTNTNHFQAKKILVILRL 233
OY 246 MEERKIDKEEDOEERKTEDEAEMSKSDHVRKQRTDLDLWVYKHS---NLSTEQI 301
DB 234 IEERR-----ASKEIQHDMGLYLMNEATRKLDDDM 266
OY 302 LDIILSLFAGHETSVAIALAIFLQACPKAVEELREHLEIARAKKEGSELMWDDY 361
DB 267 IDLIITLLSGEYVSTSMMAVKYLHDHPKYLEELRKHMAIREKKE--EDPIDYNDY 324
OY 362 KKMDFQCYINETLRGYNVRFHRAKLDVRYKGYDIPSGMKVLPVIAVHLDNSRYQ 421
DB 325 RSMKRTFRAVILETSRLATVNGVLRKTTODMEINGIYIKGRITTYTRELNDPRLPD 384
OY 422 PULFNFMWQOONGNASSGSGSFTWGNMYPFGGPRLCAGSELAKLEMAVFIHHLV 481
DB 385 PVSFNFMWMDKS-----LEHNSFLVFGGCTQCGKELGVALEISPLHYFYT 433
OY 482 KPNWELAEEDOPFAFPFVDFPNCPLIRVS 510
DB 434 KYRWEIRGGDKLMKPPRVEAPNGLRIRVS 462

RESULT 7

cytochrome P450 - Arabidopsis thaliana
N:Alternate names: protein T18B16.200; protein T5K18.10
M:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 28-Jul-2000
C:Accession: T04444; T05806
R:Byan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Bancroft, I.; Mewes, H.W.
submitted to the Protein Sequence Database, April 1998
A:Reference number: 215359
A:Accession: T04444
A:Molecule type: DNA
A:Residues: 1-457 <BEV>
A:Cross-references: EMBL:AL021667
R:Experimental source: cultivar Columbia; BAC clone T18B16
R:Byan, M.; Van Der Schuren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, April 1998
A:Reference number: 215453
A:Accession: T05806
A:Molecule type: DNA
A:Residues: 131-457 <BEW>
A:Cross-references: EMBL:AL022580
R:Experimental source: cultivar Columbia; BAC clone T5K18
C:Genetics:
A:Map position: 4
A:Introns: 67/2; 173/3; 302/3; 358/1; 393/3
A>Note: T18B16.200; T5K18.10
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreductase
P:272-433/Domain: cytochrome P450 homology <P45>

Query Match 23.6%; Score 632; DB 2; Length 457;
Best Local Similarity 31.8%; Pred. NO. 1-8e-33;
Matches 151; Conservative 92; Mismatches 176; Indels 56; Gaps 12;
OY 13 LLLPSLLSLFLILILKRNKRTRENLPFGKSGMPLGTTIGYLAFTYATTGLGDMQOHV 72

DB 10 LFRAGSLFVFLKRLISQRRFGSSKLPPLPGTMGPVVGTE---FQLYSODP-NVFPQSKO 65
OY 73 SKYGIYRNSNLFGEPTIVSADAGLNRFILIONBGRLECSYPRISGILGKMSMLV 132
DB 66 KRGSVFKTHVAGCPVMSISPEAKFVLYTSHLFKFPFPAKEMMLKQALFFHQGY 125
OY 133 HRDMRSISLNFSLHARLTKLKYERHTLVYDSMOONSIFSADEAKKFTFNMAKH 192
DB 126 HAKRLVLRAPMPEISIRN-NVPDIESIADSLRWS-EGTMTVYQEMKTYFNVALSI 183
OY 193 MSMDPGEET---EOLKKEVYFMKGVYAPNLGTAVHAKLOSATILKIEK 249
DB 184 F---GKDVLYREDIKRYLYLEKYNMVPNLPTFLHKSARKKELSOILARLSR 239
OY 250 KLDIKEDOEERKTEDEAEMSKSDHVRKQRTDLDLWVYKHS---NLSTEQI 308
DB 240 R-----QNGSSH-----NDLDSFMDKRELDEQIADNITGV 272
OY 309 LFRAGHETSVAIALAIFLQACPKAVEELREHLEIARAKKEGSELMWDDYKMDFTQ 368
DB 273 IFARPTVASVSMILKYLAEVNPVLEAVTEBQMAI-RDKERGES-LTWGDTKKMPLRS 330
OY 369 CYINETLRGYNVRFHRAKLDVRYKGYDIPSGMKVLPVIAVHLDNSRYQPNLFNM 428
DB 331 RVIQETLRVASILSTFRAVEDVEYGLIFPGMKVLPVLFNRIHSADIFSNPKFDS 390
OY 429 RMQOONGNASSGSGSFTWGNMYPFGGPRLCAGSELAKLEMAVFIHHLV 483
DB 391 RFE-----VAKRPTFMFPFGNGTSHCPNELAKLEMSIMIHLLTKY 432

RESULT 8

probable cytochrome P450 At2g29090 [imported] - Arabidopsis thaliana
N:Alternate names: cytochrome P450 homolog T914.17
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02739; D84692
R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.X.; Mason, T.M.; Shen, M.;
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.
A:Reference number: 214710
A:Accession: T02739
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-482 <ROU>
A:Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461849
R:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vent
Nature 402, 761-766, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:1061719
A:Accession: D84692
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <STO>
A:Cross-references: GB:AB002093; NID:g3461849; PID:NAC3325.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g29090; T914.17
A:Map position: 2
A:Introns: 80/2; 187/3; 328/3; 321/3; 351/3; 413/3; 453/2
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology
P:291-453/Domain: cytochrome P450 homology <P45>

Query Match 22.8%; Score 611; DB 2; Length 482;
Best Local Similarity 29.8%; Pred. NO. 4.5e-32;
Matches 153; Conservative 101; Mismatches 192; Indels 68; Gaps 13;
OY 16 PSLLSLFLILILK-----RNRKRTRENLPFGKSGMPLGTTIGYLAFTYATTGLGDMQ 69

Db 20 PALTLTIWVVVLLFKMWMKEQRLRLPPSGMLPIGFT---LALTYENP-NSFEA 75
 QY 70 QHVSATGKIYNSNLFGEPTIYSADAGLNRFILONEGRLEFESYPSSTIGIILCKMGLVLY 129
 Db 76 TRONKRGDIFFTTHILGCCVMISSEPAARMVLSKAHLFKPTYPSPKRRMIGPEALFPHQ 135
 QY 130 GDMHEDM-RSISLNLFLSHARLRTLLKDYERHTLFLVDSWOONSIFSAODEAKKTFFMLM 188
 Db 136 GPHSTLKRILVOSSFPMSALRPTV--SHIELLVLOTLSWTSOXSINPLEYMKRRAFPDVA 193
 QY 189 AKHMSMDPGEEE--TEOLKREYVTFMKGVASAPLNPCTAYHAKALOSRATILIKFIRK 245
 Db 194 ---IMSAGDKKEEPTIDIVIKILYQRLRGYNSMPLDPLGLFKHSMARLEISELAKV 250
 QY 246 MEERLIDIKEDQEEBEVKTEDEAMSKSDHYRKOROTDILLGWLY----KHSNLSREQ 300
 Db 251 IERRENGREE-----GGLIYGLVLCADQKRNGLSDSQ 283
 QY 301 ILDLISLFLFAGHETSSVALAIFELQACPRAVEELREHELET-ARAKKELGSELNWD 359
 Db 284 IADNLIGVIFPATDTTASVLTWMLKYLHDHNLQEVSRQPSIRQIKKE--NRRLSWE 341
 QY 360 DYKMDFTQCVINETLRGLNVRFLHRKALDVRKGYDIPSGWKVLPVISAHLNDSRY 419
 Db 342 DTRKMPDITRVIOETLRASVLSFTFREAVQDVEYDGLIPKGMKVLPLFRRIHSSSEFF 401
 QY 420 DQPLNFNWRMOQONNGASSSGSFSFTWGNMYPFGGPRLCAGSETAKILEMAVFIHHL 479
 Db 402 PDPEKFDPSRE-----VAPKPYTMPFGNCHVSCPSSELAKLEMLILLHL 448
 QY 480 VLKFNWELAEDDOPFAF-PFVDFPENGPIRYSRI 512
 Db 449 TTSFRWEVIGDEGIQYGFPPVKKGLPIRYTPI 482

RESULT 9

C1417
 cytochrome P450 d13695c - Arabidopsis thaliana
 N:Species: oxidoreductase (EC 1.-.-.-)
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 A:Variety: columbia
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: C71417
 R:Bevan, M.; Bancroft, I.; Bent, E.; Loyer, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dink
 P.; Medler, H.; Wedder, E.; Wambolt, R.; Weitzmeger, T.; Pohl, T.M.; Terry, N.; Gled
 avanagh, T.; Hempel, S.; Kotter, P.; Entlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
 erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Aus
 C.; Chailwitzis, N.
 A:Title: Analysis of 1.9 Mb of continuous sequence from chromosome 4 of Arabidopsis thal
 A:Reference number: A71400; MID:98121113; PMID:9461215
 A:Accession: C71417
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-487 <BEV>
 A:Cross-references: GB:Z97338; NID:g2244870; PIDN:CAB10309.1; PID:g2244888
 C:Genetics:
 A:Gene: d13695c
 A:Map position: 4C0D9-4G3845
 C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
 F:433/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match

Best Local Similarity 21.7%; Score 581; DB 1; Length 487;
 Matches 143; Conservative 100; Mismatches 205; Indels 76; Gaps 11;

QY 17 SLTSLILFLILKRRNRKTFENLPFGSGWPLGTIGYLRKYATATTGDMOAHSKY 76
 Db 3 SLPLVITFHHVYQWRNPKTKGKLPQSGMGPFRTGTFEEFKPRDALQSTFIKDRVLRFP 62
 QY 77 -----KIYRSNLFGEPTIYSADAGLNRFILONEGRLEFESYPSSTIGI-----LG 121

Db 63 ADESSILHSFRTSLFDGKALISMDMELN-----LMAKANSVPQTKSVIRLFG 112
 QY 122 KWSMVLVGDHNRDMRSLSLFLSHARLRTLLKDYERHTLFLVDSWOONSIFSAODEAK 181
 Db 113 ENNLFLOSKEHKKHVRNLTFFQLDPQGLKSMIDVDVLARTVMEGARNGVLDYKERTS 172
 QY 182 KFTFNLMKHHM-SMDPGEETEOUKREYVTFMKGVASAPLNPCTAYHAKAL----- 232
 Db 173 KILGGLAKKMGEMEP--EAKELALCKRFQSGWFRFFLNLPGYGYKMKVLPVOYT 230
 QY 233 -----OSRATILKFTERKMEERLIDIKEDQEEBEVKTEDEAMSKSDHYRKOROTDDL 286
 Db 231 EADISQARKMKMLRKRTVLTFRASGELEFFNIIJFGEKEG----- 274
 QY 287 LGWILKHSNLSREQDILDLISLFLFAGHETSSVALAIFELQACPRAVEELREHELETAR 346
 Db 275 -----GETMSVENAVEITYTFELVANETPRILAAVVKFISDHPKYOELQREHEIVR 328
 QY 347 AKKELGSELNWDYKMDFTQCVINETLRGLNVRFLHRKALDVRKGYDIPSGWKVLPVISAHLNDSRY 406
 Db 329 GKAE-KGGGLWMEYKGMHFTQWYNESLRITSTAPVLYKLEHDFQVGDITTPAGWTM 387
 QY 407 PVISAHLNDSRYDQPLNFNWRMOQONNGASSSGSFSFTWGNMYPFGGPRLCAGSE 466
 Db 388 G-YPHIHFNSEKYEDPAFNPWRMEGKDLGAIYS-----KTFIFGAGRILCYGAE 437
 QY 467 LAKLEMAVFTIHVLKFNWELAEDDOPFAF-PFVDFPENGPIRYSRI 510
 Db 438 FAKQMAVFTIHHL-FRYRMSKSGTTIIRSFMLPFGGCDVOIS 480

RESULT 10

A84859
 probable cytochrome P450 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 23-Mar-2001
 C:Accession: A84859
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bento, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
 euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MID:20083487; PMID:10617197
 A:Accession: A84859
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-485 <STO>
 A:Cross-references: GB:AEO02093; NID:g4512670; PIDN:AAD21724.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g42850
 A:Map position: 2
 C:Superfamily: Synechocystis cytochrome P450 slr0574; cytochrome P450 homology
 C:Keywords: heme; iron; metalloprotein
 F:432/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match

Best Local Similarity 21.6%; Score 579.5; DB 2; Length 485;
 Matches 147; Conservative 107; Mismatches 203; Indels 53; Gaps 11;

QY 12 LLLPLSLSLFL-----ILKRRNRKTFENLPFGSGWPLGTIGYLRKYAT-TL 64
 Db 14 LCIATWISLTFEPRKKNHRTYKIKQKKK--LLPGEMLPWIGTMDPYKAKNSRNF 71
 QY 65 GDFMOHVSRYGKIYNSNLFGEPTIYSADAGLNRFILONEGRLEFESYPSSTIGIILCKM 124
 Db 72 EDVFNRIIRKHSIFETRLMGSPITVYNGAEANRLTLSNFSILVSSWSSSSVOLMGMC 131
 QY 125 MVLVGDHNRDMRSLSLFLSHARLRTLLKDYERHTLFLVDSWOONSIFSAODEAKKT 184
 Db 132 IMAKQEKHVRNLTFFQLDPQGLKSMIDVDVLARTVMEGARNGVLDYKERTS 191
 QY 185 ENLMKHHM-SMDPGEETEOUKREYVTFMKGVASAPLNPCTAYHAKALOSRATILKEI 242

Db 192 FVVEECYLGIVETGMEV-----FERNLEGVAFALPVEFCSPFARKKRLLETFL 245
243 ERKMEERKLDIKEEDOEVEEKTEDEAEMSKSDHVRKORTDDLLGWYLKHSNSTEQL 302
246 VGVKREKREMEKEGAE-----KPNITLFSVLVELLKGV-----ITEEV 287
303 DILSLFPGHGTSSVATLALFPIQACPKAVEELREHELEAKAKKEGESE-LMWDY 361
288 DMVLLVFAHDTSTYAMSMTEKMLAQHPTCRDTLLOEHAQT---KANKGEDEYLVEV 344
362 KMDFTQCYINETLRGNVRFELHKKALKDVKYKGYDIPSGKVLPIVSAVLDSNRXYQ 421
345 KMKXSMQVVRRTMLSPFIIEGSPKRAVDIDYGGTITPKMKILMTTYGTHYNEIFPD 404
422 PVLFPWMOQONNGASSSGSFGSTWNNYMPFGGPRLCAGSELAKEMAVFIHHLVL 481
405 PMSFPTRRDKPIQAVT-----YLPFGGPRLCAGHQLAKISILVFMHVVYT 451
482 KFMWELAEODOPFAFPFDFPN-GLPIKVS 510
452 GFDWSLYVPDETISMDPLPFPGLMPKIKIS 481

RESULT 11

H86185
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86185
R:Thellogis, A.; Becker, J.R.; Palm, C.J.; Federapfel, N.A.; Kul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nucleotide 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shino, P.; Southwick, A.M.; Sun, H.; Tjallin,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H86185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <STO>
A:Cross-references: GB:AE005172; NID:g2388581; PIDN:AA071462.1; GSPDB:GN00141
C:Genetics:
A:Map position:
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 21.6%; Score 579.5; DB 2; Length 490;
Best Local Similarity 28.6%; Pred. No. 5e-30;
Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;
34 KTRFNLPGKSGMPFLEGTIGYLPYATYATLGDPMOQHVSKYK--IYRSNLFGEPTIYS 91
40 ENRHLLPFGDGLGMPFISGMLSLRAFKTSDDPSFTRILKRGPGYIKAHAFGNSPIIV 99
92 ADAGINRFLTONEGRLEFCSYPRSIGILGKMSLVLDGMDHRMRSISLNFSLHARLRT 151
100 TTSDFCRVRLTDDD-AFKPGMPTSMELIGRKSFGVIGSEEHKRLRLRLTAAPVNHGALS 158
132 ILLKVERHTLFVLDMSOONSIFSNODEAKKTFENLMAKHINSMDPGEETEQAKKEYT 211
159 TYPIETEEVIVLDKMKMGFEFLTHLRKLTFRIL--MYIFLSESESEVMDALEREVYA 217
212 FMKGVVAPLNLPGTAYHKAQSRATILKTERKMEERKLDIKEEDOEVEEKTEDEAEM 271
218 LNTGVRAAVNLPGFAHKAARCTLVAAFGSIVTER-----NQRKNTLSKKKML 271
272 SKSDHVRKORTDDLLGWYLKHSNSTEQLDILSLFAGHGTSSVATLALFPIQACP 331
272 DNLLNWK-----GKTLDEDEIIDLVLMTYLNAGHSGHTIMATVYLOEHP 320

QY 332 KAVEELREHELEIARAKKEGESELMWDYKKMDFTQCYINETLRGNVRFELHKKALKD 391
Db 321 EYLQRAKAEQOEIMLSRPE-GQKGLSIKETRKEEFISQVDETLRYTFSLAFREAKTD 379
QY 392 VRYKGYDIPSGKVLPIVSAVLDSNRXYQPLIFNWRKQOONNGASSSGSFGSTWNN 451
Db 380 VMANGSLIPKGVKVLWFRVDVHIDPEVFPDPKPFDPAR---DNG-----FVPRKA 428
QY 452 YMPFGGPRLCAGSELAKEMAVFIHHLVLFKFMWELAEODOPFAF 496
Db 429 FLFPAGSHLCGNDLAKLEISILFHLFLTKYQVRSNDECVMY 473

RESULT 12

T48973
cytochrome P450-like protein - Arabidopsis thaliana
N:Alternate names: protein P14D17.40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C:Accession: T48973
R:Jordan, N.; Banger, S.; Wiedemann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd,
submitted to the Protein Sequence Database, April 2000
A:Reference number: 225008
A:Accession: T48973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <JOR>
A:Cross-references: EMBL:AL353992; GSPDB:GN00061; ATSP:P14D17.40
C:Genetics:
A:Gene: ATSP:P14D17.40
A:Map position: 3
A:Introns: 70/2; 178/3; 229/3; 312/3; 343/3; 346/1; 381/3
C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 21.5%; Score 577; DB 2; Length 455;
Best Local Similarity 28.9%; Pred. No. 6.6e-30;
Matches 145; Conservative 84; Mismatches 173; Indels 100; Gaps 10;
32 NKRTRFNLPGKSGMPFLEGTIGYLPYATYATLGDPMOQHVSKYKRYNSLFGEPTIYS 91
28 NPKSNGKLPFGSMGFPFIIGETIDPFKPYGFYISLYLKKMLRYGLPFTNLGKYTVVS 87
92 ADAGINRFLTONEGRLEFCSYPRSIGILGKMSLVLDGMDHRMRSISLNFSLHARLRT 151
88 TDKDVNMEILROENKSFILSYDGLMKPLGKDSLFLKIGNIKHKKIQTILHLSSEGLR 147
QY 152 ILLKVERHTLFVLDMSOONSIFSNODEAKKTFENLMAKHINSMDPGEETEQAKKEYT 211
Db 148 KILKMDRYTRREHLSKATGRLDVKDAVSKLIILHLPKMS-----NLKPQTA 198
QY 212 FMKGV-----VSAPLNLPGTAYHKAQSRATILK--FIERKMEERKLDIKE 255
Db 199 KLMGIFKAFTEPWFRTSYLSAGKLYNTLW--ACREGKREIKDITYARKTEEEKY---- 252
226 EDQEEVEKTEDEAEMSKSDHVRKORTDDLLGWYLKHSN-----LSTEQILDLISLFL 310
253 -----DDFLNTAIEESEKAGELNENVAITTLTFLTSLC 284
QY 311 AGHETSSVATLALFPIQACPKAVEELREHELEIARAKKEGESELMWDYK-KMDFTQC 369
Db 285 VTQDTTSAICLAVFLLENPRVLAELKKEH-EVILSESEDEGEGVTEWEEYHKKTFTN- 342
QY 370 VINETLRGNVRFELHKKALKDVKYKGYDIPSGKVLPIVSAVLDSNRXYQPLIFNWR 429
Db 343 -----MKSGLTIPAGWLVMIIPSVYHNDPELETENPFEFNPWR 379
QY 430 WQOONNGASSSGSFGSTWNNYMPFGGPRLCAGSELAKEMAVFIHHLVLFKFMWELAE 489
Db 380 WEGKELRAGS-----KTFVFGTGLRCAGAEFARLQISVLEHILVYTYNFSLHQ 429
QY 490 DDOPFAFPFVDFPNGLPIRVSR 511

DB 430 DCEVLRVPAHLPNGISINISK 451

RESULT 13

B84733

Probable cytochrome P450 (Imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: B84733

Ridin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

enus, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84733

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1489 <STO>

A:Cross-References: GB:AE002093; NID:g3831452; PIDN:AAC69934.1; GSPDB:GN00139

C:Genetics:

A:Gene: Atg32440

A:Map position: 2

C:Superfamily: synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 20.4%; Score 547.5; DB 2; Length 489;

Best local Similarity 27.6%; Pred. No. 5,9e-28;

Matches 146; Conservative 105; Mismatches 211; Indels 67; Gaps 14;

DB 8 TLPLLLPSLLFLI-LIKRRN-----RKRPNLPKSGMPLGTTIGYLK 57

DB 4 TGLILMWF-LILGLFVLKWLKRVNVIYVSKLGEKKHYLPDGLGWMVIGNMSEFLR 62

DB 58 PYATATLGDPMQOHVSKYK--TYRSNLGEPPIVSADAGLNRFITLONEGRIFECSPRS 115

DB 63 AFTSDPESTQSYTRIGTGTYKAMFEPVLYTPPTCRRVLDLDD-AHHIGPKS 121

DB 116 IGGILKWSMLVVGDMHRMRSISLNFSLHARLITLLKVERHTFLVDSMOQNSIFS 175

DB 122 TMLIGKRSVGVISFEHKKRLRLTSAPVNGPEALSVYIGFIEVYVTDLEKMSKGEIE 181

DB 176 ADEAKKFTTNMAKHTMSMDPEEETOLKEVYTPMKVVSAPLNLPGTAHKAQSR 235

DB 182 FLSHLRKLTPEKVI-MYFISSESEHVDLSERETNLNYSVGRAMGILPGFAVHAKLAR 240

DB 236 ----ATLKRERKMERKIDI--KEEDQEEEEKTEDEAMSKSDHVRKQRTDDDLGW 289

DB 241 KTLVAFOSIVTNRNRKONISSNRKMDLNDLVDKDE-----NGRVLDLDD--- 286

DB 290 VLKHSNLSTEQIIDLILSLFAGHETSSVAIALAIFPLQACPKAVEELREHLEIARAK 349

DB 287 -----BEIIDLILMYLNACHESGCHLTMMATILMOHPHILQAKKEQBEHL--VKR 335

DB 350 ELGSEELNMDYKRMDFQCYINETLRLGNVRLHAKLADVRYKGYDIPSGKVLPIV 409

DB 336 RAGQOKLTLEKREMYLVLSQVIDETTLVITFSLTAFFREAKSDVQMDYIIPKGVLTWF 395

DB 410 SAYHLNLSRYDQPLNFMWMOQONNGASSGSGFSTWGNVWPEGGGRRLCGLSEAK 469

DB 396 RNVLHDEIYPPDKKPEPSRWE-----GYTPKAGTFLPGLSGHLLCGNDLAK 443

DB 470 LEMAVFTIHLVLKFNWELAEEDQFAFPVDFPGLP-----IRVSRLL 513

DB 444 LEISIFLHFLKRVKRSNPGCVMF-----LPHNRKDKCLAIITRTM 488

RESULT 14

D96813

hypothetical protein T30F21.17 (Imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 19-Apr-2002

C:Accession: D96813

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alou

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,

ansen, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzid

Rizzo, M.; Rooney, T.; Rowley, D.; Sakao, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D96813

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1460 <STO>

A:Cross-References: GB:AE005173; NID:g4836883; PIDN:AAD30586.1; GSPDB:GN00141

C:Genetics:

A:Gene: T30F21.17

A:Map position: 1

C:Superfamily: synechocystis cytochrome P450 slr0574; cytochrome P450 homology

Query Match 18.6%; Score 497.5; DB 2; Length 460;

Best local Similarity 26.2%; Pred. No. 9,3e-25;

Matches 133; Conservative 102; Mismatches 205; Indels 67; Gaps 12;

DB 13 LLPLSLILLFLILKRRNRKTRFNLPPKSGMPLGTTIGYLKPYATATLGDPMQOHV 72

DB 9 LMWALVAVNISHMLYKWSNPKCPKLPDSMGPIIGETLDFPKPGVGIPTFVKKRM 68

DB 73 SKYGIKTRSLNLFGEPIIVSADAGLNRFITLONEGRIFECSPRSISGGLKWSMLVVGDM 132

DB 69 IRYGPLERTNIFGSKTYVSDPDVLIHQIFQENTSEFLGPDVFKVFGDNLFLEKVF 128

DB 133 HRMRSLNLFSLHARLITLLKVERHTFLVDSMOQNSIFSADAKKFTFNLMAKH 192

DB 129 HKYLOKTRMILISSEGAKQMLGMDKATDHRISASQSFNKRKVELNVAAYMPKL 188

DB 193 MSMDPGEETEOLEKKEVYTPMKVVSAPLNLPGTAHKAQSRATILFERKKEERKLD 252

DB 189 ISNLKPEQGLKIDN-----LNANFNDWPKSF-LRLSTWKAATKALKSR--- 231

DB 253 IKEEDQEEEEKTEDEAMSKSDHVRKQRTDDDLGVLKH-----SNLSTEQIIDLILS 307

DB 232 -----EALIQWKKVDLMRKRETERKQ---EDFLNTLLELEKDGSPFDGSAINILFL 281

DB 308 LIFAGHETSSVAIALAIFPLQACPKAVEELREHLEIARAKKESEELNMDYK-KMDF 366

DB 282 LAPALREGTSSCTALAVKFTSKDPKYLAEKREKALVYDNKRD-KEAGVSNEETRHAKTF 340

DB 367 TOCVINETLRLGNVRLHAKLADVRYKGYDIPSGKVLPIVSAVHLNLSRYDQPLF 425

DB 341 TNWVSNVLRANTPTPLFRKAVQDVEIKKY-----YLEVIMHG----- 382

DB 426 NPMRMOQONNGASSGSGSSTWGN-NYMPGGGPRCAGSELAKEMAVFTIHLVLKFN 484

DB 383 --WLMQ-----GKEMWGSKTFMAFGYGVRLCGAEFSRLQMAIFLHLVAYD 429

DB 485 WELAEEDQFAFPVDFPGLPPIRVSR 511

DB 430 FSNVQDSSEIIRSPHQTOKLLINISQ 456

RESULT 15

T02263

cytochrome P450 DMARP3 - maize

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Zea mays (maize)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: T02263

R:Winkler, R.C.; Helentjaris, T.

Plant Cell 7, 1307-1317, 1995

A:Title: The maize dwarf3 gene encodes a cytochrome P450-mediated early step in gibbe

A:Reference number: Z14648; MUID:96004534; PMID:7549466

Query Match	36.5%;	Score 978.5;	DB 1;	Length 472;
Best Local Similarity	41.9%;	Pred. No. 1.7e-51;		

[illegible]

RA Van der Schueren J., Grymopriaz B., Chung Y.-J., Vandebussche F.,
RA Breken M., Welfjens I., Voet M., Bastiaens I., Aert R., Defoer E.,
RA Wiltzenegger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzner E., Brandt A., Peters S., van Staveren M., Distse W.,
RA Woolfman P., Klein Lankhorst R., Rose M., Hauf J., Koester P.,
RA Bernerstorfer S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA de Keyser A., Buysaert C., Gielen J., Villarroel R., de Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharte M., Grimm M., Loehner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fattmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argitlou A., Vitale D., Liguori R., Piravandi E.,
RA Massenat O., Hüller F., Chaboud G., Mendenhall A., Falber R.,
RA Schnabel S., Hiller R., Schmidt W., Lecherry A., Aubourg S.,
RA Chedder F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Giddons T., Weber N., Vandenberg M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Purrelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Schollier P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R. K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheel P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Hamon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Mink P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Krimer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Speleth J., Ryan E., Andrews S., Gelsel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
RA Grant E., Marra M., Martensson R., McCombie W. R.,
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana".
RL Nature 402:769-777(1999).
CC -1- FUNCTION: MIGHT BE INVOLVED IN THE BIOSYNTHESIS OF STEROIDS
CC IMPORTANT FOR THE POLAR ELONGATION OF CELLS DURING DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum
CC (Potential).
CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: AB008097; BAA37167.1; -
CC EMBL: AL161589; CAB80304.1; ALT INIT.
CC EMBL: Z99708; CAB16850.1; ALT INIT.
CC EMBL: AL022141; CA18139.1; ALT SEQ.
CC InterPro: IPR001128; Cytochrome_P450.
CC Pfam: PF00067; P450; 1.
CC DR PRINTS: PR00385; P450.
CC DR PROSITE: PS00086; CYTOCHROME_P450; 1.
CC KW Oxidoreductase; Monooxygenase; Transmembrane; Heme;
CC Endoplasmic reticulum; Multigene family.
CC TRANSMEM 4 24 POTENTIAL.
CC BINDING 463 463 HEME (BY SIMILARITY).
CC CONFLICT 45 45 F -> L (IN REF. 2).
CC SEQUENCE 524 AA; 59389 MW; 550578908BDDF272 CRC64;
Query March 32.2%: Score 864; DB 1; Length 524;
Best Local Similarity 35.4%: Pred. NO. 1.3e-44;
Matches 186; Conservative 96; Mismatches 165; Indels 78; Gaps 10;
24 FLIL-----LKRNRKTRN-----LPGKSGWFLGCTIGYL---56

```

DB 30 FLVLTAGILLRPMFLRLRSLKTRKDDDEEDNEKKGMTPNLSGLPVGELTNFAC 89
OY 57 ---KPYATATLGDPMOAHVSKRYNSNLGEPPIVSADAGINLETQNEGRLEFEC 113
DB 90 YSRPVT-----FMDKRSLGKVFETNIGPPIISTDAEVNKKVYLQHNHGFVAPR 143
OY 114 RSLGGLGKSMVLVGDHMRMSISLNTLSARLRTILKQVREHTFLVDSWOONST 173
DB 144 KSTELLGEMSLISINPQKRLHTLIGALRSPHLKDRITRDIEASVVLTLASMAOLPL 203
OY 174 FSNODEAKKFTFMKMHISMNDGPEETEQLEKEYTEKGVASAPLPGAYHIALQ 233
DB 204 VHVODELKKATFEFLKVLNLSISPG-EDMNLKLEFEETKGLCLPIKPGFRLKSLK 262
OY 234 SRATILKFERKMEERKLIDKEEDDEEEVKT-----EDEAKMSKSDHVRKQRTDDL 287
DB 263 AKRRLIKMKVKKVVEERQVAMTTSPANDVYDLVLRGDSKOSQSPEDVSGK----- 315
OY 288 GWLKHNSLSTEOILDLILSLFAGHETSSVAIALAIFLQACPKAVEELREHLEIARA 347
DB 316 -----IYEMMIGDEETMPTAMTLAVKFLSDNPVALAKLVEEMEKRR 358
OY 348 KRELSEELWMDYKKMDFQCYNIELRLGNVRFRLKALDVRKKGDISSGVLP 407
DB 359 KLELGE-EYKWTYMSLSTFQVNIETLRMANINGWRKALDVEIKGLIPRGCVLA 417
OY 408 VISAVHLNDRSYDOPNLFNPMRQOONNGASSGGSFSTWGNVYMPFGGPRCAGSEL 467
DB 418 SFSVHMDIEDIYNPQFDPWRMDRINGSANSSIC-----FTFPGGQRLCPLEL 468
OY 468 AKLEMAVFIHHLVKFENWELAEDDQPAFPFVDPNGLPPIVSRI 512
DB 469 SKLEISIFLHLVTRYSW-TAEEDLVSFPTVKMKRRLPIRVATV 512

RESULT 3
CP85_LYCES STANDARD; PRT; 464 AA.
ID CP85_LYCES 043147;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
CY Cytochrome P450 85 (EC 1.14.-.-) (Dwarf protein).
GN CYP85 OR D.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. GCR758;
RX MEDLINE=6626705; PubMed=8672892;
RA Bishop G.J., Harrison K., Jones J.G.D.;
RA "The tomato Dwarf gene isolated by heterologous transposon tagging
RT encodes the first member of a new cytochrome P450 family."
RL Plant Cell 8:959-969(1996).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U54770; AAB17070.1;
CC InterPro: IPR001128; Cytochrome_P450.
CC Pfam: PF00067; P450.1
CC PRIN2: PR00385; P450.
CC PROSITE: PS00086; CYTOCHROME_P450; 1.

```

```

KW Oxidoreductase; Monooxygenase; Membrane; Heme.
FT BINDING 414 414 HEME (By similarity).
SQ SEQUENCE 464 AA; 53706 MW; D2B21AAB7B14E94 CRC64;

Query Match 28.4%; Score 761.5; DB 1; Length 464;
Best Local Similarity 33.4%; Pred. No. 1,4e-38;
Matches 170; Conservative 90; Mismatches 188; Indels 61; Gaps 10;

OY 12 LLLPFLSLLEFLILKRRN--RKTFRNLPPGSGMPLEIGLYKPTATTLG-DFM 68
DB 5 LIFLSFPGICIFRFTALLRNNOVYKNLPPIGMGMPRLGTEFLK-----LGRSPM 58
OY 69 QOHVSKYKTYRSNLGEPPIVSADAGINLETQNEGRLEFECSPRSIGSLGKMSMLV 128
DB 59 KQOFARYGSEFSHILGCPPIVSMDSELNRYILVNEAKGLVPGQGMIDILGKCNIAV 118
OY 129 VGDMHMRMSISLNTLSARLRTILKQVREHTFLVDSWOONSTFSAODEAKFTNNLM 188
DB 119 NSAHKYMRRALLSLISPTMIROQLPRIDEFMRSHLTNW-DNKVIDIOERTKNMALSS 177
OY 189 AKHISMNDGPEET---EOLKEEYVTEPMKGVASAPLMPGAYHKKALOSRATILKTERK 245
DB 178 LKQI-----AGLESTSLAQEFMSFEFNVLGTLSPILPNTNHRGQAKIYNLRTL 233
OY 246 MEERKIDKEEDDEEEVKTDEAEKMSKSDHVRKQRTDDLGLWLNKS---NLSTEOI 301
DB 234 IEERR-----ASKEIQHMLGLVLMNEEATRFKLTDEM 266
OY 302 LDILSLFAGHETSSVAIALAIFLQACPKAVEELREHLEIARAKELGESLMDYD 361
DB 267 IDLITLTLGSGYETVSTSMNAVKYLDHDPVLEELREHMAIREKKRP-EDIDIDNY 324
OY 362 KKMDFQCYNIELRLGNVRFRLKALDVRKKGDISSGVLPVISAVHLNDRSYDQ 421
DB 325 RSMFTRAVILERSRLATYNGVLRKTTQOMEINGYIIPKGRKRYVTRRLANDPRLPD 384
OY 422 PNLFNPMRQOONNGASSGGSFSTWGNVYMPFGGPRCAGSELAKLEMAVFIHHLV 481
DB 385 PYSEFNPMRMDKS-----LEHQNPSFLVGGGTROCPGKELGVAISTFLHVFVT 433
OY 482 KFNWELAEDDQPAFPFVDPNGLPPIVSRI 510
DB 434 KYRWEELGDKLMPFVEAPNGLIRVS 462

RESULT 4
C883_ARATH STANDARD; PRT; 490 AA.
ID C883_ARATH 023051;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
CY Cytochrome P450 88A3 (EC 1.14.-.-).
GN CYP88A3 OR AT1G05160 OR YUP8H12.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gail J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,

```


RA Miltscher J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
RA Pat G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shin P., Southwick A.M.,
RA Sun H., Talign L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterbeck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.M.;
RT "Sequence and analysis of chromosome 1 of the plant *Arabidopsis*
RT *thaliana*," 816-820(2000).
RL Nature 408:816-820(2000).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AC000098; AAB71462.1; -
DR HSP6: P14779; 10P2.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
FT BINDING 6 POTENTIAL.
FT SEQUENCE 439 439 HEME (BY SIMILARITY).
SQ SEQUENCE 490 AA; 56409 MW; 7FD8CD7A8864D298 CRC64;

Query Match 21.6%; Score 579.5; DB 1; Length 490;
Best Local Similarity 28.6%; Pred. No. 1.1e-27;
Matches 133; Conservative 107; Mismatches 192; Indels 33; Gaps 9;

OY 34 KTRFNLPGKSGMPFLEGTIGYLPKPYATTLGDFMOQHSKYGK--IYRSNLFCEPTIVS 91
DB 40 EHRHVLPPQDGLWPIGNLMSFLRAFKTSDPSFTRLIKRGPGIYAHNFQNPSTIV 99
OY 92 ADAGLNRLFLONEGRLEFECSTYRSTIGLIGKMSLVLDGMHRMKSISLNFSLHARLT 151
DB 100 TTSDCRRLTDD--AFKRGMTSMELIGRKSPFGISFEHKKRLRLTAAVNGHEALS 158
OY 152 ILKQVRRHTLVLDSDMOONSIFSAODEAKKFTFNLMAKHIMSNDPGEETQLKKEYVT 211
DB 159 TITPIEENVITVLDKTKMGFEFLTHIKRLTFRIT--MTIFLSESEVMDALEREETA 217
OY 212 FMKGVVSAPLNPGTAYHKAQSRATILKFERKMEERKLDIKEEDOEVEEYKTEDEAM 271
DB 218 LNYGVRAVAVNIPGFAYHRAALKAKTLVAAPQSYTERR-----NQRKQNLNKKMDL 271
OY 272 SKSDVHAKRQDDDLGWLKHSNLTSTQDILDLISLFAGHETSSVAIALAIFLQACP 331
DB 272 DNLNWK-----DEKTLDDDEIIVLLMLYINAGHESGHITMAVTVLQEHF 320
OY 332 KAVELREHLEIARAKKELGESLMDYKMDFTQVINTLRGNVRLHRAKLD 391
DB 332 EYLQAKADEQKILKSRP--GQKGLSLKETRMKELSOVVDLTVITFSLTAERAKTD 379
OY 392 VAYKGYDIPSGKVLPIVSAVHLSRYDQPLFNPMWMOQNNAGSSGSGSFTWGN 451
DB 380 VEMNGVILPIKMGKVLWVRDVAIDPEVFPDPKFPAPRW---DNG-----FVPAKGA 428
OY 452 YMPGGGPRLCAGSELAKEMAVTTHLVLKFNWELADDDQFAF 496
DB 429 FLPPGAGSHLCPGNDLAKLEISIFLHFLKQVRSNDECPVMY 473

RESULT 5
CP26_BRAKE STANDARD: PRF; 492 AA.
AC P79739;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoic acid-metabolizing
DE cytochrome) (P450RA1) (Retinoic acid 4-hydroxylase).
GN CYP26.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Euryarchaeta: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97094702; PubMed=8939936;
RA White J.A., Guo Y.-D., Baetz K., Beckett-Jones B., Bonasoro J.,
RA Hsu K.E., Dilworth F.J., Jones G., Petkovich M.;
RT "Identification of the retinoic acid-inducible all-trans-retinoic
RT acid 4-hydroxylase".
RL J. Biol. Chem. 271:29922-29927(1996).
CC -1- FUNCTION: PLAYS A KEY ROLE IN RETINOIC ACID METABOLISM. ACTS ON
CC RETINOIDS, INCLUDING ALL-TRANS-RETINOIC ACID (RA) AND ITS
CC STEREOISOMER 9-CIS-RA. CAPABLE OF 4-HYDROXYLATION, HYDROXYLATION.
CC RESPONSIBLE FOR GENERATION OF SEVERAL HYDROXYLATED FORMS OF RA,
CC INCLUDING 4-OH-RA AND 4-OXO-RA.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- INDUCTION: BY RETINOIC ACID.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U68234; AAC60045.1; -
DR ZFIN: ZDB-GENE-990415-44; CYP26.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum
FT BINDING 438 438 HEME (BY SIMILARITY).
FT SEQUENCE 492 AA; 56281 MW; F0471435B2F30509 CRC64;

Query Match 18.8%; Score 503; DB 1; Length 492;
Best Local Similarity 27.6%; Pred. No. 3.8e-23;
Matches 144; Conservative 95; Mismatches 216; Indels 66; Gaps 14;

OY 7 HTLPLLLPLSLILLF-----ILKRRNRKTRPNLPKSGMPFLEGTIGYLP 58
DB 4 YLMTVFTCTVLPVLLFLAAVKLWEMLMIRVDNCRSPLPPTMGPLGETIQLI-- 61
OY 119 ILGKMSMLVYGDHMD-----MRSISLNTLSHARLTILKQVRRHTLVLDV--QONS 172
DB 119 ILGSDTLNVHGVQKNNKKKALMRAFSRDALEH-----YIPVIOQVKSALQEMLOQDS 172
OY 173 IFSADKAKFTFNLMAKHIMSNDPGEET--EQLKKEYVTFMKGVSAPLNPGTAYH 230
DB 173 CVLVYPERKKILMFRAMILLGFEDEQKLTQDEQLEVAFEEMIKNLFSLPIDVPSGLYR 232
OY 231 ALQSRATILKFERKMEERKLDIKEEDOEVEEYKTEDEAKMSKSHVAKQRTDDDLGV 290
DB 233 GLRAR-----NPLHSKIEENIRKRIODDNNENQKKYKDALQL--INSRSRDE----- 279
OY 291 LKHSNLSTEQIIDLILSLFAGHETSSVAIALAIFLQACPAVLEHREHLEIRAKE 350
DB 280 ----PFSIQAKKEATELLFGHETTAATSTATSLVAFGLNTEVQAKVE-----VQEKVE 331

OY 351 LG-----ESELNMDYKKMDFOCVINEELRLGNVRFELHKKALKDVRVYGYDIPSGWKV 406
 DB 332 MGRYTGKGLSMELLDQLKTTGCVIKETLRLNPPVGGFRVALKTPLELGYOIPKGMVY 391
 OY 407 PVISAVHLNDSRYDQNLFPNPMWMOONNGASSSGSFSTGMNNMPRGGGPRCAGSE 466
 DB 392 YSICDHPADVAPNKEEPOPERFM-----SKGLEDSRF--NYIPFGGSRMCVKE 442
 OY 467 LAKLENAVITLHLVLFKNNELAEDDQ-----PPAPEVDPP 502
 DB 443 FAKVLKIFLVELTCHCNWILSNGPMTKGTPTIYFVDNL 483

RESULT 6

C881_MAIZE STANDARD; PRT; 519 AA.
 AC 043246;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome P450 88A1 (EC 1.14.-.-) (DMAKF3 protein).
 GN CY88A1 OR D3
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. B73;
 RX MEDLINE=96004534; PubMed=7549486;
 RA Winkler R.G., Helentjaris T.;
 RT "The maize Dwarf3 gene encodes a cytochrome P450-mediated early step
 in gibberellin biosynthesis.";
 RL Plant Cell 7:1307-1317(1995).
 CC -1- PATHWAY: Gibberellin biosynthesis; early step.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROOTS, DEVELOPING LEAVES, THE
 CC -1- VEGETATIVE MERISTEM, AND SUSPENSION CULTURE CELLS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U32579; AAC49067.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.
 FT TRANSMEM 1 21 PORENITIAL.
 FT BINDING 466 466 HEME (BY SIMILARITY).
 SO SEQUENCE 519 AA; 57906 MW; 0F8977A02431B095 CRC64;

Query Match 18.3%; Score 491.5; DB 1; Length 519;
 Best Local Similarity 28.2%; Pred. No. 2e-22;
 Matches 129; Conservative 91; Mismatches 189; Indels 49; Gaps 11;

OY 36 RFNLPKSCMPFLGETIGTLKPTATTADPEMOQVHXYGK--IYRSNLFGEPTI-VSA 92
 DB 69 RARLPPEKSMPLVGGMAFLRAFKSPDAFTASFVRGRGTVRSFMSPTLVYTT 128
 OY 93 DAGINRRIILNEGRLFCSPYRISGIGTLGKSMVLVYGDHHRMRSSISLNFSLHARLRT 152
 DB 129 AEGCKVIALMDDA--FYTGPKATVALVGRSRVAMRYEDHRIKILYLAIPINGFDALTG 186
 OY 153 LLDVDEHRTFLVDSW-QQNSIFSADDEAKKFTFNLMAKHIMSMDPGEEETOLKKEYTA 211
 DB 187 YLPFDTRVTSIRAWADHGGSVETLELRMRFTKIIVQ-IFLGADQATTRALERSTYE 245

OY 212 FMKGVASAPLNPGRYAHKALOSRATIIKTERKMEERKLDIKE-----EDOEVEEYKT 265
 DB 246 LNYGRKAAINILNPGATYIGALARRRRLVAVLOGVLDERRARRARAGVSGGVDMDRLIEA 305
 OY 266 EDEAEKMSKDVHKKORTDDLLGLWLNKSHNLSTEOIILILISLIFAGHETSVAIALAIF 325
 DB 306 QDE-----RGRHLLDD-----EIIIVVWYLNAGHESGHTTMATV 342
 OY 326 FLQCPKRAVEELREHLEIARAKKELGSELNMDYKKMDFOCVINEELTLGNVRFELH 385
 DB 343 FLQENPDMFARAKAEQEAIMRSIPS-SQRGTLIDRFREMEYLSQVIDETLAVNISFVSF 401
 OY 386 RKALKDVRYKGYDIPSGWKVLPVISAVHLNDSRYDQNLFPNPMWMOONNGASSSGSF 445
 DB 402 RQATRDVFNKYLTPKKGKVLQWTRSVAMDPOYVDPPTKFPSPHWE-----GHSPRAGTF 456
 OY 446 STWGNNTMFGGPRPCAGSELAKLENAVITLHLVLF 483
 DB 457 -----LAFGLGARLCGNDLAKLEISVLFHFLGY 487

RESULT 7

C120_SYNT3 STANDARD; PRT; 444 AA.
 AC Q39990;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative cytochrome P450 120 (EC 1.14.-.-).
 GN CYP120 OR CYP OR SUR0574.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 RA Sugitani M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 RL DNA Res. 2:153-166(1995).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=97061201; PubMed=8905231;
 CC Kaneo T., Sato S., Kotani H., Tanaka A., Asanizu E., Nakamura Y.,
 CC Miyajima N., Hiroseawa M., Sugitani M., Sasamoto S., Kimura T.,
 CC Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nario K., Okumura S.,
 CC Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 CC Tabata S.;
 CC "Sequence analysis of the genome of the unicellular cyanobacterium
 CC Synechocystis sp. strain PCC6803. II. Sequence determination of the
 CC entire genome and assignment of potential protein-coding regions.";
 CC DNA Res. 3:109-136(1996).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: D64003; BAA10496.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
 KW Complete proteome.
 FT BINDING 391 391 HEME (BY SIMILARITY).
 SO SEQUENCE 444 AA; 50578 MW; 8F62A9EED3B54BDC CRC64;


```

Db 336 LCRSNQNDKJDMETLEBQKICGYIKETLRNLPVPGGFRFALTTPELNGQIKGNMVI 395
Qy 407 PVISAVHLDSRYDQPLFNPWMQOONNGASSSGSFTWGNMYPFGGPRICAGSE 466
Db 396 YSICDTHDVADIFTNKEEFNDRIYHPEDASRFS-----FIPFGGLASCYCKE 446
Qy 467 LAKLEMAVFTHHVLKFNWELAE-----DDQFAFPPVDFP 502
Db 447 FAKILKIFVVELARHCDWQLNGPPTMKTSPTVYPDNLIP 487

RESULT 9
C26A_HUMAN STANDARD; PRT; 512 AA.
ID C26A_HUMAN
AC Q9NR63; Q9NP41;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Cytochrome P450 26A2 (EC 1.14.-.-) (P450RA1-2) (Retinoic-acid
DE metabolizing cytochrome).
GN CYP26A2 OR P450RA12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=20300913; PubMed=10823918;
RA White J.A., Ramshaw H., Taimi M., Stangle W., Zhang A., Everingham S.,
RA Creighton S., Tam S.-P., Jones G., Petkovich M.;
RT "Identification of the human cytochrome P450, P450RA1-2, which is
RT predominantly expressed in the adult cerebellum and is responsible for
RT all-trans-retinoic acid metabolism."
RT Proc. Natl. Acad. Sci. U.S.A. 97:6403-6408(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Bourne S., Bauer C., Page K., Jones T.;
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Plays a key role in retinoic acid metabolism. Involved
CC in the specific inactivation of all-trans-RA. Responsible for
CC generation of several hydroxylated forms of RA, including 4-OH-RA,
CC 4-oxo-RA, and 18-OH-RA.
CC -1- ENZYME REGULATION: Has a preferred activity towards the following
CC substrates: all-trans-RA > 9-cis-RA > 13-cis-RA.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC similarity).
CC -1- TISSUE SPECIFICITY: Highly expressed in brain, particularly in the
CC cerebellum and pons.
CC -1- INDUCTION: By retinoic acids (RA).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; AF252297; AAF76003.1; -
CC EMBL; AC007002; AAF65576.1; -
CC MIM; 605207; -
CC InterPro: IPR001128; Cytochrome_P450.
CC Pfam: PF00067; P450_2.
CC PRINTS; PR00385; P450.
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC Oxidoreductase; Monooxygenase; Membrane; Heme; Mitosome;
CC Endoplasmic reticulum.
CC BINDING 441 441 HEME (POTENTIAL).
CC CONFLICT 68 68 O -> OKCTIRFVWVWVPO (IN REF. 2).
CC SEQUENCE 512 AA; 57512 MW; A06D1D9944E6726F CRC64;

```

```

Query Match 16.6%; Score 446; DB 1; Length 512;
Best Local Similarity 26.9%; Pred. No. 9.9e-20;
Matches 132; Conservative 86; Mismatches 219; Indels 54; Gaps 10;

Qy 14 LPLSLSLLEPLILLK-----RRKTRFNLPEKSSWPLGELTGLKPTATTL 64
Db 15 LAACLVSTLLAVSQQLRMATNRDCKLPRKSGMGPPLIGEHMLLOGSG-- 71
Qy 65 GDFMQOHSKYKGYIRSNLFGEPTIVSADAGLNPIITLONEERLEPESPRISGILGKWS 124
Db 72 --FQSSRRKKGKGVNFKTLLGRPLIRYTGAEVNRKILMGELHIVSTMPKSTRMLIGNT 129
Qy 125 MLVLVGDHNRMSISLNFSLHARLTLLDVERHTLFLVDSQON-SIFSADDEARF 183
Db 130 VSNISGDIHRKRKRKVFYSFHEALESYLPK-IGLVIDTLRAMSGHEAIVNVEAQKL 188
Qy 184 TPNLMARKIMSDGEETEDQKKRYVFMKGVSAPRLNTEGTAHKLQSRATILKTE 243
Db 189 TFRNAIRVLLFSPLEEDLGHLEFYQQFVDNVFSLPVDLPFSGYRRGIRARQILQKLE 248
Qy 244 RKMEERKLDKEEDQEEVEVTEDEAKMSKSDHVAKQRTDDLLGWLKHSLNSTEOLID 303
Db 249 KAIREK-----LQYQKGYLDALDLITSSKEHKE-----MTQELKD 288
Qy 304 LLSLFGHETSSVALAIFFIQACPRAYEELEEHLEZARAKKEG-----ESELN 357
Db 289 GTLELIFAAVYFTASASTSLIMQLKHPTVLEKLDE---LRANGILHSGGCPCEGTLR 344
Qy 358 WDDKRMDFQCVINETRLGNVNFELIRKALKDVRKYGYDIPSGKVLPAISAVHLDS 417
Db 345 LDTLSGLLYLCCYKLEVRLEFTPISGYRVLQFPELDFQIPKMSVWYSIRDTHDPAF 404
Qy 418 RYDQNLFPNRMQOONNGASSSGSFTWGNMYPFGGPRICAGSEIAKLEMAVPIH 477
Db 405 VFQDVNVDFDPDFSQ---ARSEDKGRF-----HLPFGGGVYTCGLKHLAKFLKVLAV 456
Qy 478 HLVLKFNWELA 488
Db 457 ELASTSRELA 467

RESULT 10
CP26_XENLA STANDARD; PRT; 492 AA.
ID CP26_XENLA
AC O93323;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Cytochrome P450 26 (EC 1.14.-.-) (Retinoic acid degrading enzyme
DE CYP26) (XCP26) (Retinoic acid converting enzyme) (RACE).
GN CYP26A1 OR CYP26.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99077803; PubMed=9857192;
RA Hollmann T., Chen Y., Grunz H., Pleier T.;
RT "Regionalized metabolic activity establishes boundaries of retinoic
RT acid signaling."
RL EMOB J. 17:7361-7372(1998).
CC -1- FUNCTION: Plays a key role in retinoic acid metabolism (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By
CC similarity).
CC -1- INDUCTION: By retinoic acid.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL: AF057566; AAC25158.1; -
InterPro: IPR001128; Cytochrome_P450.
Pfam: PF00067; P450; 1.
PRINTS: PR00385; P450.
PROSITE: PS00086; CYTOCHROME_P450; 1.
Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
Endoplasmic reticulum.
KW BINDING 437 437 HEME (POTENTIAL).
FT SEQUENCE 492 AA; 55459 MW; D1D4B7651BF2D3E CRC64;
SQ

Query Match 16.6%; Score 444.5; DB 1; Length 492;
Best Local Similarity 28.7%; Pred. No. 1,2e-19;
Matches 149; Conservative 85; Mismatches 222; Indels 63; Gaps 17;

7 HTLLPLLLPSLSLLLT-----ILKRRNRKTRNLPKSGWPFLETTIGLKP 58
4 YTLTSLALCTLALPILLTLLTAALMEVYCLRRKADACANPLPGTGLPFGFTLQMVQ 63
59 YVATITLGDMMOQVSKYKGIYNSNLFGEPTIVSADAGLAFIIONGRLEPCYPSISG 118
64 RRR-----FLQYKRSQYGRITKHLGSPTRVATGAEVNRQILMGHKLVSVMPSVPT 118
119 ILGKMSLVLVGDHMRDMSISLNFSLHARLTILKDYERHTFLVDSMOONS---TF 174
119 ILGAGCLSLHNEHKYTKVIAQAFSRALAN-YYPQHEEYRCVNLMLDSGCVLVY 177
175 SAODEKATFTFLMAKHMSMDP---GEEETDLKREYVTFMGVSAPLNPGTAYHKA 231
178 PA---TKRMKFRAMRLILGCDPQRMDRQETLLLEAFEMSRLFLSDIVPFSGLYRG 234
232 LOSRATILFIERKMERLIDKEDEEVEKTEDEAMSKSDHAKRTDDLLGWL 291
235 LRRR---NLHNAQIE---NTEKLRQRPDEHKDQLD-LIDSR--RNGEPIVQLAL 284
292 KHSNLTEDILDLISLIFAGHETSSVATAIAFFLQACPKAVEELREHLE---TARAK 348
285 KSS--ATE-----LFGGHGTTASATLSLFLAKKDLVLEKRAKE-LERQGLSLTR 333
349 KELGESELNWDYKKMDFQCYINETLRLGNVYRFLHRAKLDVRYKYGDIDSGMKVLPV 408
334 PE-EKKELEISIEVLQDLKTSYKIVKTELRLSPVAGGFVALKTFVNGYQIRKGMVYIS 392
409 ISAVHLDNSRYQPNLFNFMWQOQNNNGASSSGSSFSWGNMNYMFFGGPRICAGSELA 468
393 IADTHGEADLFPTDKFNDFRFLPLPRDSSR-----RGLPFGGGVRCICIGKEPA 443
469 KLEMAVFIHLVLFKFMEL-----AEDDPFAFPFVDFP 502
444 KILKMFVVELCRNCDWELNLSPAATSPITICPVNDLP 482

RESULT 11

CP26_CHICK STANDARD; PRT; 492 AA.
AC 09PUB4: 09PUG2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 26 (EC 1.14.--) (Retinoid acid degrading enzyme CYP26).
GN CYP26A1 OR CYP26.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic spinal cord;
RX MEDLINE=20057772; PubMed=10588879;
RA Swidell E.C., Thaller C., Sockanathan S., Petkovich M., Jessell T.M., Eichele G.
RT "Complementary domains of retinoid acid production and degradation in the early chick embryo."
RL Dev. Biol. 216:282-296(1999).
RN [2]

RP SEQUENCE OF 368-474 FROM N.A.
RA Martinez-Ceballos E., Burdall C.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: Plays a key role in retinoid acid metabolism. Appears to be involved in the specific inactivation of all-trans-RA. Responsible for generation of several hydroxylated forms of RA, including 4-OH-RA, 4-oxo-RA, and 18-OH-RA.

-1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum. -1- DEVELOPMENTAL STAGE: Expressed at stage 4 in the ectoderm, stage 5-7 in the nascent notochord and at stage 7 its expression decreases in the anterior part of the embryo. From stage 7-10 its expression is restricted to the dorsal folds of the neural tube and to rhombomere 2. At stage 10, it is expressed in the lateral plate endoderm and in the tail bud and by stage 11/12 it disappears in the neural tube, followed by a confined expression at stage 12 to dorsal neural tube and at stage 15 an increasing expression in the ectoderm.

-1- INDUCTION: By retinoid acid.

-1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL: AF199462; AAF09250.1; -
EMBL: AF185266; AAD56546.1; -
InterPro: IPR001128; Cytochrome_P450.
Pfam: PF00067; P450; 2.
PRINTS: PR00385; P450.

PROSITE: PS00086; CYTOCHROME_P450; 1.
Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
Endoplasmic reticulum.
KW BINDING 437 437 HEME (POTENTIAL).
FT CONFLICT 371 371 I -> V (IN REF. 2).
FT CONFLICT 472 474 GPI -> SPT (IN REF. 2).

FT SEQUENCE 492 AA; 55264 MW; 7F28B72E75C2322FB CRC64;
SQ

Query Match 16.4%; Score 439.5; DB 1; Length 492;
Best Local Similarity 26.3%; Pred. No. 2,3e-19;
Matches 141; Conservative 84; Mismatches 210; Indels 101; Gaps 15;

9 LRLPLLLPSLSLLLT-----RRNKRTRFNLPKSGWPFLETTIGLKP 60
6 LVASALCTFLPLLEFLAVALMDLYCAGSRDPCSLPLPGTGLPFGFTLQMVQ 60
61 ATTLDGMOHNSKYKGIYNSNLFGEPTIVSADAGLAFIIONGRLEPCYPSISG 120
61 VLRKREFLDMKRRKGFITKHLFGAPRYVAGAEVNRHILGHNLYSVQMPASVPTIL 120
121 GRKMSLVLVGDHMRDMSISLNFSLHARLTILKDYERHTFLVDSMOONSIFSAQ--- 177
121 GSGCLSNLHNGHKKRYIMQAFS-----RDALGHYVPID--EVSACLAQWIG 169
178 -----DEAKKTFPFLMAKHMSMDPGE---EETDLKREYVTFMGVSAPLNPGT 226
170 AGPCLLVPEVRLMRIRMLILGFQPOASPDGQQLVEAFEEIIRMLFSLPDIYPS 229
227 AVHKLQSRATILKFERMEERKLDIKEDEEVEKTEDEAMSKSDHAKRTDDLL 286
230 GLYGLRRR---NTHAKIEE-----NIRAKMARKE---PEGYKDA 265

```

QY 287 LGWLKHS-----NLSTEOIIDLILSLFAGHETSSVAIALAIFPLQCPRAVELREEH 341
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 266 LQLMHEHTGNGEOLNMDELKESATELLEFGHETASATSLIFLG-----HH 315
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 342 LEIARAKEL-----GESELMNDYKKMDPTOCVINETRLGNVAFRLHRAKALD 391
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 316 DVLQKREKLEOLKGLISPNOKELNMFELQKLTGCVIKETRLSPVPGGFRILAKT 375
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 392 VRYGVDIPSGWKVLPVISAVALHNSRYDQPLNPNWMOOQNGASSSGSSTWGN 451
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 376 LELNGYQIPKGMVNYISICDTHDAVLEFTDKDEFNPDFFM-----SPSPEDSRRS 426
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 452 YMPGGGPRCAGSELKLEMAVFIHHVLKFNWELAEDDQ-----PPAPFPVDP 502
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 427 FIPGGGLRSCVGRKFAKVLKLTFTVELARSCDMLNGPPTMKTPGYEVDNLP 482
| : : : : | : : : : | : : : : | : : : : | : : : : |

RESULT 12
CP26_HUMAN STANDARD; PRT; 497 AA.
ID 043174;
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 26 (EC 1.14.--) (retinoic acid-metabolizing
cytochrome) (P450RA1) (hp450RA1) (retinoic acid 4-hydroxylase).
GN CYP26A1 OR CYP26.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373542; PubMed=9228017;
RA White J.A., Beckett-Jones B., Guo Y.-D., Dilworth F.J., Bonasoro J.,
RJ Jones G., Petkovich M.;
RT "DNA cloning of human retinoic acid-metabolizing enzyme (hp450RA1)
identifies a novel family of cytochromes P450."
RL J. Biol. Chem. 272:18538-18541(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98380037; PubMed=9716180;
RA Somneveid E., van den Brink C.E., van der Leede B.M., Schulkes R.K.,
RJ Petkovich M., van der Burg B., van der Saag P.T.;
RT "Human retinoic acid (RA) 4-hydroxylase (CYP26) is highly specific for
all-trans-RA and can be induced through RA receptors in human breast
and colon carcinoma cells."
RL Cell Growth Differ. 9:629-637(1998).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=99045433; PubMed=9826557;
RA Trofimova-Griffin M.E., Juchau M.R.;
RT "Expression of cytochrome P450RA1 (CYP26) in human fetal hepatic and
cephalic tissues."
RL Biochem. Biophys. Res. Commun. 252:487-491(1998).
CC -1- FUNCTION: Plays a key role in retinoic acid metabolism. Acts on
retinoids, including all-trans-retinoic acid (RA) and its
stereoisomer 9-cis-RA. Capable of both 4-hydroxylation and 18-
hydroxylation. Responsible for generation of several hydroxylated
forms of RA, including 4-OH-RA, 4-oxo-RA, AND 18-OH-RA.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- TISSUE SPECIFICITY: Highest levels in adult liver, heart,
pituitary gland, adrenal gland, placenta and regions of the brain.
CC -1- INDUCTION: By retinoic acid.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: AF005418; AAB88881.1;
DR HSSP: P14779; 1JPZ.
DR Genew: HGNC:2603; CYP26A1.
DR MIM: 602239;
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum.
FT BINDING 442 442 HEME (POTENTIAL).
SQ SEQUENCE 497 AA; 56162 MW; EAB684B24B2EAB3 CRC64;

Query Match 16.3%; Score 438; DB 1; Length 497;
Best Local Similarity 27.0%; Pred. No. 2.9e-19;
Matches 142; Conservative 85; Mismatches 226; Indels 72; Gaps 15;

QY 10 LPLLLPSL-----LSLLFLILK-----RRNRKTRFMLPFGKSGMPLGETIGYK 57
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 3 LPALLASALCTFVLPILFLFAIKLMDLYCSGRDSCALPLPGTMGFFGELL----- 58
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 58 PYTATLGDPEMQQVSKYKTIYRSNLGEPPIVSADAGLNFTIQNEGRLEFCSPNSIG 117
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 59 -QWVLQRRKFLQMKRRKGYFTKTHLFGRPVWAGADVNRILLGDRLVSVHPASVR 117
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 118 GILKWSMLVYVGDHMRMSISNLFSLHARLR---TILKDVVERHNLFLVDSMQ--NS 172
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 118 TILSSGCLSNLHDSHOKRKVYKRAFSREALDECVPIVTEVSS--LEOMISGGER 173
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 173 IFSADKAKKPTFMIAKHMIMSDP-----GEETEQLKKEYVFMGVASAPLNPETA 227
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 174 GLIYVEYKRLMFRIARHILLCFEPQLAGDSE-QQLVEAFEMETRLNLSLPIDVFPSS 232
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 228 YHKLQSAATLTKTERKMERKLDIKED-----QEEYKVTDEADMSKSHVKKORT 282
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 233 LYRMKRAKMLHARIEINIRKIGLASEAGQCKALQILLIHSWE-----KGRRL 285
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 283 DDDLGLWKLHNSLSTEOIIDLILSLFAGHETSSVAIALAIFPLQCPRAVELREEH 342
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 286 DMQ--ALKQS--STE-----LFGGHETASATSLITLGLGYPRVLOVREBLK 331
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 343 EIAARAKELGESELMNDYKKMDPTOCVINETRLGNVAFRLHRAKALDYKGYDIPSG 402
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 332 SKGLICKSNQNKRLDMETLEQLKYIKETRLNPNVPGGFVALKTFELNGYQIPKG 391
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 403 WKVLPVISAVALHNSRYDQPLNPNWMOOQNGASSSGSSTWGNYPGGGPRIC 462
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 392 WNVYISICDTHDAVLEFTNKEEFNPDFFMPLPHPDASRRS-----FIPGGGLRSC 442
| : : : : | : : : : | : : : : | : : : : | : : : : |
QY 463 AGSELAKLEMAVFIHHVLKFNWELAE-----DDQFAPFPVDP 502
| : : : : | : : : : | : : : : | : : : : | : : : : |
DB 443 VGRKFAKILKLTFTVELARHCDMLNGPPTMKTPGYEVDNLP 487
| : : : : | : : : : | : : : : | : : : : | : : : : |

RESULT 13
YRY2_CAEEL STANDARD; PRT; 520 AA.
ID YRY2_CAEEL
AC 027514;
DI 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DI 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative cytochrome P450 CYP13A5 (EC 1.14.--).
GN CYP13A5 OR T10B9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Gardner A.;

```

RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: Z48717; CAAB8604.1; -
 DR HSSP: P14779; LJPZ.
 DR WormPep: T10B9.2; CE01656.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
 KW BINDING 464 464 HEME (BY SIMILARITY).
 FT
 SQ SEQUENCE 520 AA; 59524 MW; 087B19E25B7ADF3B CRC64;

Query Match 12.1%; Score 324.5; DB 1; Length 520;
 Best Local Similarity 23.6%; Pred. No. 1.7e-12;
 Matches 120; Conservative 94; Mismatches 178; Indels 117; Gaps 20;

42 GKSGMPFLG-----ETIGYIKPYTATTTGDFMOQHVSKYKITYRSLFGEPTVSAD 93
 33 GPRGPFVGVHIEFDYENRGLK-----LGEWKE-----YGPITIGTEGKILYVSN 82
 94 AGLNRPILQNEGRLEECSTPRISIGILG-----KWSMLVLV-GDMHRMRSISLNFSLA 147
 83 P---EFVHEVFVKQDNFGRKTNPICGDPNKKRAHLVSAQGHMKRLRTLSPEFSNK 139
 148 RRTIL-----LKVYERTILFVLDSMOQNSIFSNODEAKKTFNLMKHMSPGEE 201
 140 NLRKIMSTVEETVELMRH---LDQMSAKGKAVDLDIYQETLIDIGRIAM---GOTE 192
 202 TEOLKREYVTFMKGV-----VSAPLNLPGTAVHK-----ALQSRATILKF 241
 193 SLMFNPMLPKYKGLFKOSKRLPLVSGIFPIAGTMRFFEMRPSIQAPFIMSTVEKA 252
 242 IERKAEERKLDIKE-----EDQ-----EEBEVKTEDBAEKSQSDHYAK 279
 253 LNKRLRLEORADEKAGIEPSGEPQDFIDLFLDARAVNDFEESALGFAKTEIAKYD---K 309
 280 QNTDDDLGWLVAHNSLSTEQILDLILSLFAGHETSSVAIALAFILOACKPAAVELE 339
 310 QLFPELILG-----OLPVLLAGYDPTALISLSSYSLARHPEIQKLOE 354
 340 EHLEIARAKKEELGESELNMDYKKMDFTQCIVINETIRLGNVVRFLH-RKALDVAHYKGD 398
 355 E-----VDRECPNEVFVFDQISKLYMECVKALRYPLASLYHNKRMKMETVNLGQV 408
 399 IFSGKVLPLVSAVHLDSRY-DQPNLFNPMWMOQOONNGASSGSGSFSTWGNNTMPFGG 457
 409 IEKGIVNVQVDWTLTHYDKRWGEDANERFERWE-----SGDELFYAKG-GYLPFGM 459
 458 GPRCAGSELAKEMAVFIHILYLFKNWE 486
 460 GPRICIGMKLAMEKMKMLTHILKRYTEE 488

DE Putative cytochrome P450 CYP13A7 (EC 1.14.-.-).
 GN CYP13A7 OR T10B9.10.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Gardner A.;
 RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: Z48717; CAAB8609.1; -
 DR HSSP: P14779; LJPZ.
 DR WormPep: T10B9.10; CE01655.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.
 KW BINDING 464 464 HEME (BY SIMILARITY).
 FT
 SQ SEQUENCE 518 AA; 58999 MW; CC04283EF87B9EA7 CRC64;

Query Match 12.0%; Score 321.5; DB 1; Length 518;
 Best Local Similarity 24.0%; Pred. No. 2.6e-12;
 Matches 120; Conservative 96; Mismatches 184; Indels 101; Gaps 20;

42 GKSGMPFLG-----ETIGYIKPYTATTTGDFMOQHVSKYKITYRSLFGEPTVSAD 93
 33 GPRGPFVGVHIEFDYENRGLK-----PSEWKKKGPVYIGTEGKILYISD 82
 94 AGLNRPILQNEGRLEECSTPRISIGILG-----KWSMLVLV-GDMHRMRSISLNFSLA 147
 83 P---EFVHEVFVKQDNFGRKTNPICGDPNKKRAHLVSAQGHMKRLRTLSPEFSNK 139
 148 RRTIL-----LKVYERTILFVLDSMOQNSIFSNODEAKKTFNLMKHMSPGEE 201
 140 NLRKIMSTVEETVELMRH---LDQMSAKGKAVDLDIYQETLIDIGRIAM---GOTE 192
 202 TEOLKREYVTFMKGV-----VSAPLNLPGTAVHK-----ALQSRATILKF 241
 193 SLMFNPMLPKYKGLFKOSKRLPLVSGIFPIAGTMRFFEMRPSIQAPFIMSTVEKA 252
 242 IERKAEERKLDIKE-----EDQ-----EEBEVKTEDBAEKSQSDHYAK 279
 253 LNKRLRLEORADEKAGIEPSGEPQDFIDLFLDARAVNDFEESALGFAKTEIAKYD---K 309
 280 QNTDDDLGWLVAHNSLSTEQILDLILSLFAGHETSSVAIALAFILOACKPAAVELE 339
 310 QLFPELILG-----OLPVLLAGYDPTALISLSSYSLARHPEIQKLOE 354
 340 EHLEIARAKKEELGESELNMDYKKMDFTQCIVINETIRLGNVVRFLH-RKALDVAHYKGD 398
 355 E-----VDRECPNEVFVFDQISKLYMECVKALRYPLASLYHNKRMKMETVNLGQV 408
 399 IFSGKVLPLVSAVHLDSRY-DQPNLFNPMWMOQOONNGASSGSGSFSTWGNNTMPFGG 457
 409 IEKGIVNVQVDWTLTHYDKRWGEDANERFERWE-----SGDELFYAKG-GYLPFGM 459
 458 GPRCAGSELAKEMAVFIHILYLFKNWE 486
 460 GPRICIGMKLAMEKMKMLTHILKRYTEE 488

DB 468 BLAMMEMKMLNLIKNTFE 488

RESULT 15
 ID C912_ARATH STANDARD: PRU: 500 AA.
 AC 065790; 09S2U3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Cytochrome P450 91A2 (EC 1.14.-.-).
 GN CYP91A2 OR AT4G37430 OR F6G17.80.
 OS Arabidopsis thaliana (mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_taxid=3702;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Seedling;
 RX MEDLINE=98281573; PubMed=9620263;
 RA Micolanti M., Ward E., Ohta D.;
 RT "Cytochrome P450 superfamily in Arabidopsis thaliana: isolation of
 RT cDNAs, differential expression, and RFLP mapping of multiple
 RL plant Mol. Biol. 37:39-52(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Ansoerger W., Brandt P., Glivelli L.A., Rieger M.,
 RA Weichselbacher M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis B., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Biham L., Robben J.,
 RA Van der Schueren J., Grynoprez B., Chang Y.-J., Vandenbusche F.,
 RA Braeken M., Welljens I., Voet M., Bastlaens I., Aert R., Defoor E.,
 RA Meltzenger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirksen W.,
 RA Koolman P., Klein lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Bernerster S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail A., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIay K., Mayes R.,
 RA Petter A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loecherer T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Hertzl A.,
 RA Neumann S., Argirion A., Vitale D., Liguori R., Piravandi E.,
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 RA Schindl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Cheifons T., Weber N., Berger C., Monfort A., Casachuberta E.,
 RA Perez-Perez A., Punelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijman D., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
 RA Zaecaria P., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Parnell L., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Sekhon M., Murray J., Sheet P., Schutz K., Huang E., Spiegel L.,
 RA Stenking T., Kallio J., Graves T., Harmon G., Edwards J.,
 RA Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Iamur B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hamed A., Lodhi M., Johnson A.,
 RA Chen E., Maira W., Martienssen R., McCombie W.R.;

RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D76607; BA28539.1; -
 DR EMBL: AL035601; CAB38210.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Membrane; Heme.
 FT BINDING 438 438 HEME (BY SIMILARITY).
 FT CONFLICT 106 106 A -> V (IN REF. 1).
 FT CONFLICT 127 127 I -> M (IN REF. 1).
 FT CONFLICT 140 140 N -> I (IN REF. 1).
 FT CONFLICT 454 454 S -> T (IN REF. 1).
 SQ SEQUENCE 500 AA; 57555 MW; 0FBA53D2070EA2EA CRC64;

Query Match 11.88; Score 316; DB 1; Length 500;
 Best Local Similarity 25.34; Pred. No. 5,4e-12;
 Matches 135; Conservative 83; Mismatches 199; Indels 118; Gaps 22;

QY 9 LRLPLPLPSSLLPLFLILKLR--RNKRPENLPDSCSGMPGEGITGIAKPTATILGD 66
 DB 1 MLRYILLP-----LLELVISYKFLYKSTQRFNLPDGPSPFVGH-DLHKRP--PIHR 50
 QY 67 PQOQVSKYKIKYKRNLFGEPTV-----SADAGLNRFLIONGRIFECSPYSIG 117
 DB 51 LLOYSNOYGPPIF-SLRFSGRRVYVITPSLAQSFQGNIDYVLSNPQLDTAKYAYNH 109
 QY 118 GILGKMSMLVGDGMHDMRSI-SLNLSHARKLT--LKDVERHTLVLDSSMOQNSIF 174
 DB 110 TVYV-----TAPYGDHMLNRLRISQETLSHRLINQHIRKIDILNLTLSRYTOTS-- 163
 QY 175 SAODEAKKTF-----MIMAKHIMSDPG-----DEETPOLKKEVVTMKGV 216
 DB 164 ---NESNDHFHIEPLPLSLTFENNIVRYMTGKRYGGDYVNNKEAELEFK----- 211
 QY 217 VSAPLNLPCTAYHKAQSRAT-----IKFTERKMERKLDIKEDDEEVEKTEDER 269
 DB 212 -----LVYDIAMYSGANHSADYLPILKLGKNEKFEVKAIG-----KMDDI 253
 QY 270 EMSKSDHVRKQRTDDDLGVNLKHSNLTBQILD-----LISLRLAGHETSVAIALAI 324
 DB 254 LQALIDECRDKGKNTVMNHLISLQOOQPRYYVDYIIGKIMSMMLAGETSSAVLTENMAM 313
 QY 325 FFLQACRAVEELREHLELTARAKKEGSEL-NMDDYKMKDPTQCVINETLRLGNVVP 383
 DB 314 ANLLRNPVEVLEKARSE-----IDEXIGDRDLDESDIANVPLQNVSETRFLFVYAPF 367
 QY 384 L-HRKALKVRYRGYDIPSGKYLPIYSAVHLDNSRYDQNPENPFWOQOONASSSGS 442
 DB 368 LIPRSPDMDKIGGYVDPRTIVNANAIHRDPEIWEEREPENPDY---NDGC----- 419
 QY 443 GSFTSTNNY-----NPFGGPRLCAGSELAKEMAVFTHHVLKRNWELAEDDQ 492
 DB 420 -----GSDIYKYMMPFGNGRTCTGAGLGGATVTLGSLQCEWENVKGEE 468

Search completed: March 30, 2003, 12:05:09
 Job time : 39 secs

...

...

...

.

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2003, 11:49:16 ; Search time 63 Seconds

(without alignments)
1677.812 Million cell updates/sec

Title: US-09-502-426A-2

Sequence: 1 MFEHTHTLLPULLLPSSLIS.....FAFPVDPNGUPINVRIL 513

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirts:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2681	100.0	513	10	064989	064989 arabidopsis
2	2677	99.9	513	10	09SC09	09SC09 arabidopsis
3	951	35.5	474	10	09LKH7	09LKH7 vigna radia
4	866	32.3	512	10	09FX29	09FX29 arabidopsis
5	843.5	31.5	491	10	094IA6	094IA6 arabidopsis
6	822	30.7	478	10	09LNT3	09LNT3 arabidopsis
7	815	30.4	382	10	09LX89	09LX89 arabidopsis
8	813.5	30.3	490	10	094IW5	094IW5 oryza sativ
9	789	29.4	464	10	09LIC5	09LIC5 arabidopsis
10	756.5	28.2	481	10	08W0R4	08W0R4 sorghum bic
11	726	27.1	465	10	09FMA5	09FMA5 arabidopsis
12	712	26.6	465	10	094OV4	094OV4 arabidopsis
13	711	26.5	465	10	09LH81	09LH81 arabidopsis
14	672.5	25.1	467	10	0949P1	0949P1 arabidopsis
15	658	24.5	463	10	09PH76	09PH76 arabidopsis
16	632	23.6	457	10	065624	065624 arabidopsis

17	631	23.5	518	10	09FI38	09FI38 arabidopsis
18	631	23.5	735	10	09LGI7	09LGI7 arabidopsis
19	622	23.2	443	10	09LJK2	09LJK2 arabidopsis
20	611	22.8	482	10	081077	081077 arabidopsis
21	606	22.6	479	10	08VZC2	08VZC2 arabidopsis
22	597.5	22.3	485	10	08W4T9	08W4T9 arabidopsis
23	596.5	22.2	497	10	09AXM6	09AXM6 taxus cuspi
24	589	22.0	486	10	09LIF9	09LIF9 arabidopsis
25	581	21.7	487	10	023384	023384 arabidopsis
26	581	21.7	496	10	09FOY4	09FOY4 cucurbita m
27	579.5	21.6	474	10	09SHY7	09SHY7 arabidopsis
28	579.5	21.6	485	10	09SHT2	09SHT2 arabidopsis
29	577	21.5	455	10	09LKH8	09LKH8 arabidopsis
30	576.5	21.5	490	10	09C5Y3	09C5Y3 arabidopsis
31	575.5	21.5	464	10	004949	004949 arabidopsis
32	569.5	21.2	477	10	09LVT7	09LVT7 arabidopsis
33	547.5	20.4	489	10	09ZV72	09ZV72 arabidopsis
34	546.5	20.4	444	10	08S917	08S917 arabidopsis
35	546.5	20.4	489	10	09C5Y2	09C5Y2 arabidopsis
36	518	19.3	504	10	09SNG3	09SNG3 oryza sativ
37	502	18.7	499	10	09AXH9	09AXH9 hordeum vul
38	497.5	18.6	460	10	09SYN2	09SYN2 arabidopsis
39	453	16.9	497	11	08VIL0	08VIL0 rattus norv
40	435	16.2	224	10	09SDM6	09SDM6 helianthus
41	421	15.7	394	10	08WIS6	08WIS6 zea mays su
42	414	15.4	394	10	08WIT1	08WIT1 zea mays su
43	414	15.4	394	10	08WIS9	08WIS9 zea mays su
44	413	15.4	394	10	08WVJ5	08WVJ5 zea mays su
45	412	15.4	394	10	08WIV5	08WIV5 zea mays su

ALIGNMENTS

RESULT 1	
064989	PRELIMINARY; PRT; 513 AA.
ID 064989	
AC 064989;	
DT 01-AUG-1998 (TREMBLrel. 07, Created)	
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)	
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)	
DE Steroid 22-alpha-hydroxylase.	
GN DWF4.	
OS Arabidopsis thaliana (Mouse-ear cress).	
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;	
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.	
OX NCBI_TaxID=3702;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=MS-2;	
RX MEDLINE=98158690; PubMed=9490746;	
RA Choe S., Dikes B.P., Fujioka S., Takatsuto S., Sakurai A.,	
RA Feldmann K.A.;	
RT "The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates	
RT multiple 22alpha-hydroxylation steps in brassinosteroid	
RT biosynthesis."	
RT Plant Cell 10:231-243(1998).	
RL -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.	
CC EMBL: AF044216; AAC05093.1; -.	
DR InterPro: IPR001128; Cytochrome_P450.	
DR Pfam: PF00067; P450; 1.	
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.	
KW Heme; Monooxygenase; Oxidoreductase.	
SO SEQUENCE 513 AA; 58867 MW; B1639BDD9A4DA6F3 CRC64;	
QY Query Match 100.0%; Score 2681; DB 10; Length 513;	
Best Local Similarity 100.0%; Pred. No. 1.6e-184;	
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DB 1 MFEHTHTLLPULLLPSSLISLFLILKRRNRKTRNLPFGKSGWFFLGTIGYLRPT 60	
1 MFEHTHTLLPULLLPSSLISLFLILKRRNRKTRNLPFGKSGWFFLGTIGYLRPT 60	

```

OY 61 ATTLDGFMQOHXSKYKGYRSMLEFGEPTIVSADAGLNRTILONEGRLEFCSTPRISIGTL 120
DB 61 ATTLDGFMQOHXSKYKGYRSMLEFGEPTIVSADAGLNRTILONEGRLEFCSTPRISIGTL 120
OY 121 GKMSMLVTVGDHMRMSISLNFSLHARLRTILKDVREHTLFLVDSMQOONSIFSAODEA 180
DB 121 GKMSMLVTVGDHMRMSISLNFSLHARLRTILKDVREHTLFLVDSMQOONSIFSAODEA 180
OY 181 KKFETNLAKHITMSDPCGEETEOLKKEVYPMKVSAPLNPGTAHKAQOSRATILK 240
DB 181 KKFETNLAKHITMSDPCGEETEOLKKEVYPMKVSAPLNPGTAHKAQOSRATILK 240
OY 241 FIERMEERKLDIKEDDEEVEVKTEDEAKMSKSHVRKQRTDDLLGVLKHSNLSTEQ 300
DB 241 FIERMEERKLDIKEDDEEVEVKTEDEAKMSKSHVRKQRTDDLLGVLKHSNLSTEQ 300
OY 301 ILDLTSLFLFAGHETSSVAIALAIFFLQACPKAVELEBEHEIARAKKEGESELNMD 360
DB 301 ILDLTSLFLFAGHETSSVAIALAIFFLQACPKAVELEBEHEIARAKKEGESELNMD 360
OY 361 YKKMDFTCVINETRLRGVNVAFELHRRKALKDVRKGYDIPSGWKVLPVISAHLNDSRYD 420
DB 361 YKKMDFTCVINETRLRGVNVAFELHRRKALKDVRKGYDIPSGWKVLPVISAHLNDSRYD 420
OY 421 QPNLEPNRMQOONNGASSSGSSTGWNNTMPRGGPRCLAGSELAKLEMAVFIHHLY 480
DB 421 QPNLEPNRMQOONNGASSSGSSTGWNNTMPRGGPRCLAGSELAKLEMAVFIHHLY 480
OY 481 LKFNNELAEDDQPFAPFPVDFPNGLPPIRYSRTL 513
DB 481 LKFNNELAEDDQPFAPFPVDFPNGLPPIRYSRTL 513

RESULT 2
O9SCQ9 PRELIMINARY; PRT; 513 AA.
AC O9SCQ9:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Steroid 22-alpha-hydroxylase (DMP4) (A13950660/T3A5_40).
GN T3A5.40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloecher H., Mewes H.W., Lemcke K., Mayer K.F.X., Quetier F.,
RA Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamlya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shin P., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,

```

```

RA Hayashizaki Y., Ishida J., Jones T., Kamlya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AL132979; CAB62435.1; -.
DR EMBL: AF412114; AAL06567.1; -.
DR EMBL: AY090266; AAL90927.1; -.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450.1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KM Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 513 AA; 58667 MW; B1639DD9A5D7C93 CRC64;

Query Match 99.9%; Score 2677; DB 10; Length 513;
Best Local Similarity 99.8%; Pred. No. 3e-184;
Matches 512; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MEFEHHTLPLRLPLSLSTLFLILKRRNRKTRFNLPEKSGWPLGRTIGLKPYT 60
DB 1 MEFEHHTLPLRLPLSLSTLFLILKRRNRKTRFNLPEKSGWPLGRTIGLKPYT 60
OY 61 ATTLDGFMQOHXSKYKGYRSMLEFGEPTIVSADAGLNRTILONEGRLEFCSTPRISIGTL 120
DB 61 ATTLDGFMQOHXSKYKGYRSMLEFGEPTIVSADAGLNRTILONEGRLEFCSTPRISIGTL 120
OY 121 GKMSMLVTVGDHMRMSISLNFSLHARLRTILKDVREHTLFLVDSMQOONSIFSAODEA 180
DB 121 GKMSMLVTVGDHMRMSISLNFSLHARLRTILKDVREHTLFLVDSMQOONSIFSAODEA 180
OY 181 KKFETNLAKHITMSDPCGEETEOLKKEVYPMKVSAPLNPGTAHKAQOSRATILK 240
DB 181 KKFETNLAKHITMSDPCGEETEOLKKEVYPMKVSAPLNPGTAHKAQOSRATILK 240
OY 241 FIERMEERKLDIKEDDEEVEVKTEDEAKMSKSHVRKQRTDDLLGVLKHSNLSTEQ 300
DB 241 FIERMEERKLDIKEDDEEVEVKTEDEAKMSKSHVRKQRTDDLLGVLKHSNLSTEQ 300
OY 301 ILDLTSLFLFAGHETSSVAIALAIFFLQACPKAVELEBEHEIARAKKEGESELNMD 360
DB 301 ILDLTSLFLFAGHETSSVAIALAIFFLQACPKAVELEBEHEIARAKKEGESELNMD 360
OY 361 YKKMDFTCVINETRLRGVNVAFELHRRKALKDVRKGYDIPSGWKVLPVISAHLNDSRYD 420
DB 361 YKKMDFTCVINETRLRGVNVAFELHRRKALKDVRKGYDIPSGWKVLPVISAHLNDSRYD 420
OY 421 QPNLEPNRMQOONNGASSSGSSTGWNNTMPRGGPRCLAGSELAKLEMAVFIHHLY 480
DB 421 QPNLEPNRMQOONNGASSSGSSTGWNNTMPRGGPRCLAGSELAKLEMAVFIHHLY 480
OY 481 LKFNNELAEDDQPFAPFPVDFPNGLPPIRYSRTL 513
DB 481 LKFNNELAEDDQPFAPFPVDFPNGLPPIRYSRTL 513

RESULT 3
O9LKH7 PRELIMINARY; PRT; 474 AA.
AC O9LKH7:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cytochrome P450.
GN CYP90A2.
OS Vigna radiata.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Fabales; Fabaceae; Papilionoideae; Phaseolae; Vigna.

```


RC STRAIN-CV. COLUMBIA;
 RA Shinada Y.;
 RT "P450 gene repressed by brassinosteroid.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AB066286; BAB62109.1; -.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE: PS00030; RNA_RNP_1; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 491 AA; 56153 MM; 02FB908A91995A40 CRC64;
 SQ
 Query Match 31.5%; Score 843.5; DB 10; Length 491;
 Best Local Similarity 37.0%; Pred. No. 1.7e-52;
 Matches 176; Conservative 104; Mismatches 157; Indels 39; Gaps 9;
 QY 40 PRGKSGMPLEGTIGYLYKRYTATTLGDEMQHVSQYKRYRNLGCEPTIVSADAGLNRF 99
 DB 52 PRGSGMPLEGTIGYLYKRYTATTLGDEMQHVSQYKRYRNLGCEPTIVSADAGLNRF 111
 QY 100 ILQNEGRLEFECSPRSIGGILGKMSMLVYGDHMDRSISINFLSHARLTILKDYER 159
 DB 112 VLQSDSTAFVPPRYKTVRELKMSKSLILNGSLHRRFHLVGSFLKPLKAOIYRDMK 171
 QY 160 HTLEVLDSSQONSIFSADDEAKKFTFNMAKIHMSMDPEEETDOLKKEVTFMGVSA 219
 DB 172 FLSESDMSEQPYLLQDVSKTVAFKYLAKALISYKQ-EDLEELKREFENFISGLM 230
 QY 220 PLNPGTAAHAKLQSRATLTKFERKMEERKLDIKEODEEVEVTEDEAE-SSKSH 276
 DB 231 PLNPFQTLHRSLOAKKNVQYERIEKIR-KTKNEEDVIAKDVVDVLLDSSSH 288
 QY 277 VKRQRTDDDLGKMSLSTEQILDLISLFGHETSVAIALAIFLOACRAVE 336
 DB 289 -----LTH-NLIANNMIMMI-----PGHDSVPVLLITLAVFLDSPAALNL 329
 QY 337 LREHELEIARAKKEGSESLANDYKKDDFTQCIVNETRLCNVYFELRKALKDYRK 396
 DB 330 LREHELEIARAKKEGSESLANDYKKDDFTQCIVNETRLCNVYFELRKALKDYRK 396
 QY 397 YDIPSGMKVLPYSAVHLNDSRYDQNLNPFPMQOONNASSSGSSTWGNVMPRG 456
 DB 369 YDIPSGMKVLPYSAVHLNDSRYDQNLNPFPMQOONNASSSGSSTWGNVMPRG 456
 QY 457 GPRPLAGSELAKLEMAVFIHLVLKFWMLAEDDQPAFPVDPNGLPIRVSR 512
 DB 437 GQORLCPGLDLARLETSVFLHVLVTRFRW-LAEDDTIINFVHMKKLPVWIKRI 491
 RESULT 6
 O9LNT3 PRELIMINARY; PRT; 478 AA.
 ID O9LNT3;
 AC O9LNT3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE T1C24.27.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OK NCBI_TaxID=3702;
 RN SEQUENCE FROM N.A.
 RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
 RA Shinn P., Altafi H., Bel Q., Chin C., Chlou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Muharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;

RT "Genomic sequence for Arabidopsis thaliana BAC T1C24 from chromosome
 RT I.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AC025417; AAF8087.1; -.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 2.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR Heme; Monooxygenase; Oxidoreductase.
 KW SEQUENCE 478 AA; 55058 MM; F67A9B9F2P2D8558 CRC64;
 SQ
 Query Match 30.7%; Score 822; DB 10; Length 478;
 Best Local Similarity 35.6%; Pred. No. 5.6e-51;
 Matches 181; Conservative 95; Mismatches 187; Indels 46; Gaps 8;
 QY 12 LLLPLSLILFLILKRRNRKTRNLPGRSGMPLEGTIGYLYKRYATTLGDMOQH 71
 DB 4 LLIMVSLILSTHMYVSRNKRCKGKLPGRSGMPLEGTIGYLYKRYATTLGDMOQH 63
 QY 72 VSKYKRYNSNLFGEPTIVSADAGLNRFILQNEGRLEFECSPRSIGGILGKMSMLVYGD 131
 DB 64 VKKYGPIFTNLVGRVLYSTADSLSYFVFNQEGRCFQSWYPTDFTLHIGKKNVSLHGF 123
 QY 132 MKRDKSISLNLSHARLTILKDYERHTLEVLDSSQONSIFSADDEAKKFTFNMAKH 191
 DB 124 MYKYLKNMYLTIFGIDGLK-KLPQVEMTANKRLLEIWSNDSVEILKDATASMIFFLTAK 182
 QY 192 INSMDEEETDOLKKEVTFMGVSAPLNPGTAAHAKLQSRATLTKFERKMEERK 251
 DB 183 LISHDP-DKSSNLNANFAAFYQGLISFPFDIPGTAHYHCLQGRKAKAMLMNMQERR- 240
 QY 252 DIKEEDQEEVEVTEDEAEKSDHYRKQRTD-DLLGKMSLSTEQI-LDLISL 308
 DB 241 -----ENPRKNPSDFEYVIEIQEGITLLEIALDLMEVL 277
 QY 309 LEAGHETSVAIALAIFLOACRAVELEHELEIARAKKEGSESLANDYKKDDFTQ 368
 DB 278 LEAGHETSVAIALAIFLOACRAVELEHELEIARAKKEGSESLANDYKKDDFTQ 368
 QY 369 CVINETLRLGNVYFELRKALKDYRK-----GYDIPSGMKVLPYSAVHLNDSRYDQ 422
 DB 337 CVINETLRLGNVYFELRKALKDYRK-----GYDIPSGMKVLPYSAVHLNDSRYDQ 396
 QY 423 NLNPFPMQOONNASSSGSSTWGNVMPRGGPRPLAGSELAKLEMAVFIHLVLK 482
 DB 397 LVFNPSRME-----GSKVTNASKHFAFGGMRCVGTDTFKLQMAAFSLHLYTK 446
 QY 483 FWMLEAEDDQPAFPVDPNGLPIRVSR 511
 DB 447 YRMEELKGNITRTGLODPNGYHVKLHK 475
 RESULT 7
 O9LY89 PRELIMINARY; PRT; 382 AA.
 ID O9LY89;
 AC O9LY89;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical 43.9 kDa protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OK NCBI_TaxID=3702;
 RN SEQUENCE FROM N.A.
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
 RA Rudd S., Lemke K., Mayer K.F.X.;
 RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 [2]

RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project.
 RC Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AL163817; CAB87779.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; p450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN.1.
 KW Heme: Hypothetical protein; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 382 AA; 43869 MM; IBC5685AB73B4E30 CRC64;

Query Match 30.4%; Score 815; DB 10; Length 382;
 Best Local Similarity 34.2%; Pred. No. 1.3e-50;
 Matches 159; Conservative 87; Mismatches 127; Indels 92; Gaps 4;

QY 45 GPEPGETIGYKPTATTLGDPMOOHVSKGKIYRSNLFGEPTIVSADAGINREITLONE 104
 DB 2 GPEPGETISFEPKPHRSISIGFTLQORVSRGKFKVSNCGGKAVVSCDQELNMFILONE 61
 QY 105 GRLFCSTPRISIGILGKMSLVGDMHDMRSISLNPISHARLRLTILDKVENHTLFV 164
 DB 62 GRLFTSDVPKAMHDLGKYSILANGELIHRKILKNYISITLTKSKPDLHCAENLSISI 121
 QY 165 LDSMOQNSIFSAQDEAKFTFNLAHKHMSMDPGEETEOUKKEVYTMKGVSAPLNP 224
 DB 122 LKSMNCREVEERKEVKFTFLSVYVNOQLSTIKPEDPARLYLQDPLSMKGPISLPLP 181
 QY 225 GYAVKALQSRATILKFERKMEERKLDIKEDDEEVEYKTEDEAEMSKSDHVRKQRTDD 284
 DB 182 GTGYNALKRVSNRNHONATIEDNNNAIREDFDLSITSNED----- 224
 QY 285 DLGAVNLKHSNLSLEQILDLILSLFAGHETSSVALAIFLQACRAVELREHLEI 344
 DB 225 -----EERAAI 230
 QY 345 ARAKKEGESELMNDYKAKMDFTOCYINETLRGNVAFELRKALKDVRKYGYDIPSGWK 404
 DB 231 -PAKKGDEL-LNMEEDYOKMEFTOCYISEALRCGNITVYTRAKHDIKNEYIYIPKWK 288
 QY 405 VLPVYSAVHLNDRDQPNLEPNWRMOQNNNGASSSGSFTSGNNMPRGCGPRCAG 464
 DB 289 VFPTFVHLDPSLHENDFEENPRKWTKT-----AGGGVAVRCGP 329
 QY 465 SELAKEMAVFIHHLVLFKEMLEADDOPAFPVDPNGPLIRV 509
 DB 330 GELCKLQIAFPLHLVLSYRWKIKSDMPAHNPVEFKRGLLEI 374

RESULT 8

ID 0941W5 PRELIMINARY; PRT; 490 AA.
 AC 0941W5;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Cytochrome P450-like protein.
 GN P0419B01.5.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriharoidae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone: P0419B01."
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: A0003244; BAB56089.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; p450; 1.

DR PROSITE: PS00086; CYTOCHROME P450; UNKNOWN.1.
 KW Heme: Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 490 AA; 54824 MM; 9EC263BBAFB88F CRC64;

Query Match 30.3%; Score 813.5; DB 10; Length 490;
 Best Local Similarity 34.2%; Pred. No. 2.4e-50;
 Matches 173; Conservative 106; Mismatches 156; Indels 71; Gaps 9;

QY 27 LKRRNR-----KRRNLPKSGMPFLGETIGYKPTATTLGDPMOOHVSKYK 77
 DB 28 LKRRRRAGSGKGDAAAARLPQSGFVPGVGETLEFVSCAYSRPREAFDKRRKLHGS 87
 QY 78 -LYRNLGEPPTVSADAGINRFILQNGRLFECSYPRSIGIGKMSMLVYGDHMD 136
 DB 88 AVFRRHLSGATVYADAEVRFVLSARAFVWYPRPSLTELKSKSLILINALQRRV 147
 QY 137 RSIISNFSARLRTILKDYERHTLFVLDWQNSIFSAQDEAKKFTFNLAHKHMSMD 196
 DB 148 HGLVGAFFKSSHLKSQILDADMRRLSPALSSFPSSSLHVOHLAKSVFEITLVGLGLE 207
 QY 197 PGEETEOUKKEVYTMKGVSAPLNPJGTAVHRAQSRATILKFERKMEERK----- 250
 DB 208 AG-DEMQLKQFOEFYIGLMSLPKLPETRLYSLOAKKRMARLIQRIREKARRAA 266
 QY 251 -----LDKEEDDEEVEYKTEDEAEMSKSDHVRKQRTDDLLGAVNLKHSNLSLEQILDL 304
 DB 267 SPRDAIDIVLIGDSDE-----LTDLISDN 292
 QY 305 ILSLFAGHETSSVALAIFLQACRAVELREHLEIARAKKEGESELMNDYKMK 364
 DB 293 MIDMLPAEDSVPLITLAVKFLSECPALHOLEEENIQLRKRTDMGET-LQTDVMSL 351
 QY 365 DFTOCVINETLRGNVAFELRKALKDVRKYGYDIPSGMYLPVYSAVHLNDRYDQPNL 424
 DB 352 SFTGHVTELRGNIIIGIMKRAVRYEKGHILIPGKCFYFRSVHLDVLTLDYDPYK 411
 QY 425 FNPWRMOQNNNGASSSGSFTSGNNMPRGCGPRCAGSELAKEMAVFIHHLVLFKN 484
 DB 412 FNPWRMEKD-----MNGSFT-----PFGGGRICPGILDARLEASIFLHLVTSFR 459
 QY 485 WELAEDDOPAFPVDPNGPLIRVS 510
 DB 460 W-VAEEDHYNFPVRLKRGMPIRVT 484

RESULT 9

ID 09LICS PRELIMINARY; PRT; 464 AA.
 AC 09LICS;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Cytochrome P450-like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE-20363099; PubMed-10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 Sequence features of the regions of 4,251,695 bp covered by ninety pl,
 TAG and BAC clones."
 RL DNA Res. 7:217-221(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL: AP001307; BAB01922.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00067; P450.1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 DR PROSITE: PS00030; RNA_RNP_1; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 464 AA; 52864 MW; A05E094665E5190B CRC64;

Query Match 29.4%; Score 789; DB 10; Length 464;
 Best Local Similarity 36.5%; Pred. No. 1,3e-48;
 Matches 163; Conservative 101; Mismatches 145; Indels 38; Gaps 8;

QY 40 PRKSGKGPPLGTTGYLKYATTTAGDEMOQHVSYKTIKNSLFGTEPTVSADAGLNEF 99
 DB 52 PHSLSGMPVIGETIEFVSASDREPSFMDKRRRLMYGRFKSHIGTATYSTDAEVNNA 111
 QY 100 ILONEGRLEFECYSPRSIGGILGKMSMLVYSGDMHRDMSISLNFSLHRLRTILLKDYER 159
 DB 112 VLSQSTAVFPPTVRELMLGKSSILLNGSLHRRFGLVSGSLKPLKAOIVADMHK 171
 QY 160 HTLFVLDSMOQNSIFSADDEAKFTFNLMKAIMSDPGEETEOLEKREYTFMKGVSA 219
 DB 172 FLSEMDLMSDEQPVLLDDVSKTAVFKYLAALISVEKG-EDLEELKREFEFISGLMSL 230
 QY 220 PLNLPGTAVHKAQSRATLTKFERKMERKIDIKEDQEEBEVTEDEAEN---SKSDH 276
 DB 231 PINFGTOLHRSLOAKKNVKKQVERIIIEGKIR--KTKNKEEDVDIAKDVLKDSSEH 288
 QY 277 VRKQTDDDLGGVLAHNSLSTEOILLDLISLFGHEHSSVAIALAFFLQACRKYAE 336
 DB 289 -----LTH-NLIANNMIDMI---PGHDSVPLILTLAKFSLDSAPALNL 329
 QY 337 LREHELEIARAKKEGESELEMDYKKMDFTQCVINETLRLGNVRFELRKALDVRKYG 396
 DB 330 LEEEMMKL-KSLKELTGEPLVNDYLSLPTQKVTETILRKQNVILGYRKMKMVEIKG 388
 QY 337 YDIEGKVLVPLVYSAVHLDSNRYDQNLFPWRMOQONNGASSSGSFSFTGNMYPFG 456
 DB 389 YVIPGWCFLYLRSVHLDKLYESPRYKFNPRMQRDMNTSS-----FSPEG 436
 QY 457 GGPRLCAGSELAKLEMAVEFHHLYLKF 483
 DB 437 GGQRLCPGLDLAKLETSVFLHHLVTRF 463

RESULT 10
 O9MORA PRELIMINARY: PRT: 481 AA.

AC O9MORA:
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative cytochrome P450-like protein.
 GN SB32H17.4
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Sorghum.
 OX NCBI_TaxID=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV; BTX623;
 RA Park Y.-J., Ramakrishna W., Sanmiguell P., Emberton J., Bennetzen J.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV; BTX623;
 RA Liaca V., Young S., Kovchok S., Messing J.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF466201; AAL73972.1; -
 DR InterPro: IPR001128; Cytochrome_P450.

PK Pfam: PF00067; P450.1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 SQ SEQUENCE 481 AA; 54984 MW; 6145CEPFOFD74FE CRC64;

Query Match 28.2%; Score 756.5; DB 10; Length 481;
 Best Local Similarity 33.5%; Pred. No. 2.9e-46;
 Matches 170; Conservative 104; Mismatches 190; Indels 43; Gaps 7;

QY 15 LPSLSLLEFILLK-----RRNRKTRPNLPKSGMPFLGTTGYLKYATTTAG 65
 DB 6 LAALSVTLGAILLRMAFKMNMVGRGREGMLPPSGSGFLPETLEFFAASPTLLEV 65
 QY 66 DEMQHVSYKTIKNSLFGTEPTVSADAGLNRFLQNGRLFECSYSPRSIGGILGKSM 125
 DB 66 PEFKRLEFEGPIFRNTNIGEDMIVSLDELMARVLOOEENGFQIWTYSSFMRIIGADNM 125
 QY 126 LVLVQDMHRDMSISLNFSLHRLRTILLKDYERHTLFVLDSMOQNSIFSADDEAKFTF 185
 DB 126 VSMGLPRLHRIINLVRLFGPEALNLVLLRDVORSARDELRSWLDREVEYVTTATSRMTF 185
 QY 186 NLMKAIMSDPGEETEOLEKREYTFMKGVSAPLNPGTAVHKAQSRATLTKFERK 245
 DB 186 GTYAKKLISHD-DVAGSGSLMKCFDAMTKGLSPICVPGTAFYCGMGKRMVKVLMQO 244
 QY 246 MEERLIDKEDEQEEBEVTEDEAKSKSDHYRKQRTDDDLGVLKHSNLSTEOI-IDL 304
 DB 245 IDERR-----NGAERTYDFDL-----VIDELDKPMSIMSEIALNL 282
 QY 305 ILSLFAGHEHSSVAIALAFFLQACRKYAEELREHELEIARAKKEGESELEMDYKKM 364
 DB 283 LFLFLFASHETTSKGLTYLKLTLNPKSLQELTEHEHEMERRVD-PPSDITWEIYSKM 341
 QY 365 DETQCVINETLRLGNVRFELRKALDVRKYGYPDPSGKVLVYSAVHLDSRYDQNL 424
 DB 342 KTSVHIESRLANLAPVFRQADVHIKGYTIPESKIMICPSAHLNSKYVEDPLA 401
 QY 425 FNPWRMOQONNGASSSGSFSFTGNMYPFGCGGRLCAGSLALMAVFIHHLYLKF 484
 DB 402 FNPWRMKTDPPEVGS-----KDFMAFGGLRLCVGAEFAKLQAMFLHYLVTRF 452
 QY 485 WELAEDDQFAPFPVDFPGLPIRYSR 511
 DB 453 WKALSKGTMMLYPLGLRFPDGFHIQLHK 479

RESULT 11
 O9FMAS PRELIMINARY: PRT: 465 AA.

AC O9FMAS:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Cytochrome P450 (Brassinosteroid-6-oxidase).
 GN BROOXI.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=98290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:41-54(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=21295570; PubMed=11402205;

RA Shinada Y., Fujioaka S., Miyauchi N., Kushihiro M., Takatsuto S.,
 RA Nomura T., Yokota T., Kamiya Y., Bishop G., Yoshida S.,
 RT "Brassinosteroid-6-oxidases from Arabidopsis and tomato catalyze
 RT multiple C-6 oxidations in brassinosteroid biosynthesis.",
 RL Plant Physiol. 126:770-779(2001).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AB009048: BAB0653.1; -
 DR EMBL: AB035868: BAB0653.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 465 AA; 53767 MW; 0C00459C0866D1F CRC64;
 Query Match 27.1%; Score 726; DB 10; Length 465;
 Best Local Similarity 32.3%; Pred. No. 4.3e-44;
 Matches 166; Conservative 91; Mismatches 181; Indels 76; Gaps 10;
 QY 12 LLLPSSLFLFLILKRRN--KTRRRLPRGKSGWPLGTTIGLYKPYATTLGDFMQ 69
 DB 11 LLIIVSLCSALL-----RNMQRRTKNGLPGLGMPJFGSTTFELKGP-----NFMK 59
 QY 70 QHVSRYGKIYRNLGEPETIVSADAGLNRFILQNGRLFECSYPSIGIGIKMSMLV 129
 DB 60 NQRLRGSEFKSHLIGCPILISMDESVNRKILAKNSKGLVPGYPSMDLIGTCMAAVH 119
 QY 130 GDMHDMSSISLNTSHARKLRTILKDYERHTLFYLDSSMOONSISADDEAKKFTF 185
 DB 120 GSHRLMKSLSLISSTPMRDHILPKYDHEFRSYLDQMLEVLDIDDKTKHMAFLSL 179
 QY 186 -----NLAKHIMSDPGEETEOLEKKEVYTFMGVAPLNPCTAYHAKLOSATTLK 240
 DB 180 TOIANLKKRPV-----EETKAFKLVGTLVSPIDLPCTNRCGIQANNND 229
 QY 241 FIERMEERKLDKEEDOEKEEVKTEDEAEMSKSDHVRKORTDDLLGWVLKHSN---L 296
 DB 230 LLRELMQERR-----DSGE-----TFDDMGYLKKKEGNRP 262
 QY 297 STEOILDLISLPAGHETSSVAIALAIFLQACPKAVEELREHLEIARAKKEGESEL 356
 DB 263 TDELRDQVVTLLSGYETVSTSMALIKYLHDKALQELRAELAFERERKQ--DEPL 320
 QY 357 NMDYKKMDTQCVINETLRIGNVRFELHRRKALDVRKGYDIPSGVPLVISAIVHND 416
 DB 321 GLEDVSKMFTFRAVYETSRILATVNGLRKTTDLENGLRILKGRRIYVYTEINDA 380
 QY 417 SRVDPNLFNFWRMOQONNGASSSGSSESTWGNMYPFGGPRCLAGSELAKLEMAVFI 476
 DB 381 NLYEDPLLFNFWRMKKSLSESO-----NSCEVFGGGRCLCPKEIGIVEISFL 429
 QY 477 HHLVLFKFMELAEDDOPFAFPVDFPNGLPIRVS 510
 DB 430 HYETFRYRMEIGDELMAVFPFVAPAGFHLRIS 463
 RESULT 12
 Q940V4 PRELIMINARY; PRT; 465 AA.
 AC Q940V4;
 DT 01-DEC-2001 (TREMREL. 19, Created)
 DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
 DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
 DE A0330180/T20F20_6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.-J., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carinici P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,

RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.,
 RT "Arabidopsis cDNA clones";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kim C.-J., Chen H., Cheuk R., Meyers M.C., Banh J., Bowser L.,
 RA Carinici P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Shinozaki K., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,
 RT "Arabidopsis ORF clones";
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AY063728; AA36078.1; -
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 465 AA; 53814 MW; 280A21D0712PCA11 CRC64;
 Query Match 26.6%; Score 712; DB 10; Length 465;
 Best Local Similarity 32.1%; Pred. No. 4.3e-43;
 Matches 163; Conservative 94; Mismatches 194; Indels 56; Gaps 8;
 QY 14 LLLPSSLFLFLIL-----KRRNRKTRFLPRGKSGWPLGTTIGLYKPYATTLGDF 67
 DB 3 IMMILGLIVIVLCIRALLRNQMRYSKGLPRTGMPJFGSTTFELKGP-----DF 57
 QY 68 MOHVSRYGKIYRNLGEPETIVSADAGLNRFILQNGRLFECSYPSIGIGIKMSMLV 127
 DB 58 MNQRLRGSEFKSHLIGCPILISMDESVNRKILAKNSKGLVPGYPSMDLIGTCMAAVH 117
 QY 128 LVGDHDMSSISLNTSHARKLRTILKDYERHTLFYLDSSMOONSISADDEAKKFTF 187
 DB 118 VHGSHRLMKSLSLISPTMMKDHLPKIDDFERNYLCMDLETVIOEKTKHMAFL 176
 QY 188 MAKHIMSMDPGEETEOLEKKEVYTFMGVAPLNPCTAYHAKLOSATTLKFEKME 247
 DB 177 SSLQIAETLKKPEVEYRTEFEFLVGLISVPIIDIGTYRBSGVCAARNIDRLLELMQ 236
 QY 248 ERKLDKEEDOEKEEVKTEDEAEMSKSDHVRKORTDDLLGWVLKHSN---LSTQI 303
 DB 237 ERK-----ESGEPT-----DMIGYLMKKEDNRYLLTDEKIRD 269
 QY 304 LILSLPAGHETSSVAIALAIFLQACPKAVEELREHLEIARAKKEGESELNMDYK 363
 DB 270 QVVTLLSGYETVSTSMALIKYLHDKALQELRAELAFERERKQ--DEPLTDDIS 327
 QY 364 MDTQCVINETLRIGNVRFELHRRKALDVRKGYDIPSGVPLVISAIVHND 416
 DB 328 MKFTFRAVYETSRILATVNGLRKTTDLENGLRILKGRRIYVYTEINDA 387
 QY 424 LFPNFWMOQONNGASSSGSSESTWGNMYPFGGPRCLAGSELAKLEMAVFIHHLVLF 483
 DB 388 IFNFWRMKKSLSESO-----YFLTFGGVRLCPKEIGISEVSSFLHYETKY 436
 QY 484 NWEIAEDDOPFAFPVDFPNGLPIRVS 510
 DB 437 RWEENGDKLMAVFPFVAPAGFHLKCS 463
 RESULT 13
 Q9LH81 PRELIMINARY; PRT; 465 AA.
 AC Q9LH81;
 Q9LH81;

01-OCT-2000 (TREMBlrel. 15, Created)
 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 Cytochrome P450.
 Arabidopsis thaliana (Mouse-ear cress).
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Kosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA:
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA:
 RA MEDLINE=20363099; PubMed=10907853;
 RX Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 CC -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL, AP002060, BAB02270.1, -.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PS00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 465 AA; 53862 MW; 630A21D0765E0DD CRC64;

Query Match 26.5%; Score 711; DB 10; Length 465;
 Best Local Similarity 32.1%; Pred. No. 5,1e-43;
 Matches 163; Conservative 93; Mismatches 195; Indels 56; Gaps 8;

14 LLLPILLSLLFLIL-----KRRNRKTRFNLPPKSGMPFLGRTIGYIKPTATLGD 67
 3 IMMILLGLLVITVICLTALLRNMRKSKGLPPTMGMPJFGTTEFLKQGP-----DE 57
 68 MOOHVSKYKTIYRSNLFCEPTIVSADAGLNRFILQNEGRLEFCSYPSRIGTIGKMSMLV 127
 58 MKNQGLRARGSPFKSHILGCPITVSMDAELNRYILNENSGKLVAGYPSQMLDILGTCNTAA 117
 128 LVGNHRDMRSISLNFSLHARLITLKDVERHTLFLDSWQNSIFSAODEAKKFTNL 187
 118 VHGPHRLMRGSLSLISPTMKDHLPKIDFMNRYLCGMDELTVDIQEKTKHMAF-L 176
 186 MAKHMSMDPGEETEOAKKEVTFMKGVASPLNPGTAYHKALOSATILKTERKME 247
 177 SSLQIAETLKKPEVEERYRTEFFKLVGTLVSPIDIPGTNRSFGFQARNNDRLTLELMQ 236
 248 EKKLIDKEEDEEEBEVKTEDEAKMSKSDHVRKQRTDDLLGVNLKHSN-----LSTEQILD 303
 237 ERK-----ESGETFL-----DMLGVLKMKEDNRYILYLTKEIRD 269
 304 LILSLFPGHETSSVAIALAIFLQACRAVEELREHLEIARAKKELGESELMNDYKK 363
 270 QVVTLISGYEIVSTSMALKYLLHDHKALEELREHIAIRERRP--DEPLTLDIDKS 327
 364 MDETCVINEFLRLGNVYRFLHRKALKDVRKGYDIPSGMKVLPYISAVHLDNSYDDPN 423
 368 MFTFAVIFETSRILATYINGVLRKTHDELNGVLLIPKGMRYVYVTRREINDTSLYEDPM 387
 424 LFNPRMOQNGASSSGSGSEFTGNVMPGSPRLCAGSELAKLEMAVYIHLVLKF 483
 388 IENPRMKEKLESKS-----YFLFEGGVRLCPGKELGISVSSLHIFVYKY 436
 484 NMELAEDDQPFAPFVDPNGLPIRVS 510
 437 RWEENGEDKLWFPFVSAKPGYHLCSS 463

RESULT 14
 ID 0949P1 PRELIMINARY; PRT; 467 AA.
 AC 0949P1;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative cytochrome P450 protein.
 GN ATG19230.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Kosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A.
 RP Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
 RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
 RA Carlncl P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Kosema E., Lam B.,
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full Length cDNA of gene ATG19230 (GI:7268718).";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE FROM N.A.
 RP Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carlncl P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamuya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;
 RT "Arabidopsis Open Reading Frame (ORF) Clones.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL, AY050980; AAK3657.1; -.
 DR EMBL, AY091446; AAM14385.1; -.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam, PF00067; P450; 1.
 DR PROSITE, PS00086; CYTOCHROME_P450; UNKNOWN_1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 467 AA; 53037 MW; 2F24230446536D955 CRC64;

Query Match 25.1%; Score 672.5; DB 10; Length 467;
 Best Local Similarity 32.3%; Pred. No. 3e-40;
 Matches 164; Conservative 96; Mismatches 183; Indels 65; Gaps 15;

13 LLLPILLSLLFLILKRRNRKTRFNLPPKSGMPFLGRTIGYIKPTATLGDPMQHV 72
 10 LEAGSLFLYFLRCLLSQGRFSSKILPPTMGMPYVEET---PQLYSQDP-NVFFQSKQ 65
 73 SKYGTIYSNLFCEPTIYSDAGLNRFILQNEGRLEFCSYPSRIGTIGKMSMLVYGD 132
 66 KRYGVFETHVYLGCPVVISPEAKFVLYTSHLFTFPFASKRMKQKQAIFFPHOGY 125
 133 HDMSISLNFSLHARLITLKDVERHTLFLDSWQNSIFSAODEAKKFTNLMAKI 192
 126 HAKLKLVLRAFMPESTRN-NVPDLESTADSLRSM-BETMNTYQEMKTYFFNALLSI 183
 193 MSMDPEEET---EQLKREYVTFMGVVSAPLNPGTAYHKALOSATILKTERKME 249
 184 F---GKDEYLRDLKRCYITLLEKGYNSMPVNLGTLFHKSMARKKELSOILAILISR 239
 230 KLDIKEEDDEEVEKTEDEAKMSKSDHVRKQRTDDLLG-VWLKHSNLTQEDIDLLISL 308
 240 R-----ONGSSH-----NDLISFGDKDEEELTDEQIADNIIGV 272
 309 LEAGHETSSVAIALAIFLQACRAVEELREHLEIARAKKELGESELMNDYKKMDFTQ 368


```

Db 273 IFARDTASVMSWILKIYLAENPNVLEAITEOMA-I-RKDEKES-LTWGDTKKMPLTS 330
QY 369 CYINETLRGNVRELHRRALDVKYKGYDIPSGMKVLPVISAHLNDSRYDQPLNPMW 428
Db 331 RVIIOETLRVASILSTFTRAEVDEYEGYLLIPKGVLPFRNTHSHSDITSNPKETDPS 390
QY 429 RMOOONNGASSSGSFSFTGNNYMPFGGPRLCAGSELAKLEMAVFTIHLVLFKNWL- 487
Db 391 RE-----VAPKPNTEFMPFGNGTHSCPGNELAKLEMSIMTHLTKYSMSIV 437
QY 488 -AEDD---QPAFPVDPNGLPIRVSR 511
Db 438 GASDGIQYGFALP-----QNGLPYALAR 461

RESULT 15
Q9FH76 PRELIMINARY; PRT; 463 AA.
AC Q9FH76;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cytochrome P450 (AT5g45340/K9E15_12).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
NCBI_TaxID=3702;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asanizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty pl and YAC
RT clones.";
RL DNA Res. 7:31-63(2000).
RN 12;
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
RA Becker J.R.;
RT "Arabidopsis cDNA clones."
RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -1-SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB020744; BAB10255.1; -
DR EMBL; AY065065; AAL57698.1; -
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
KV Heme; Monooxygenase; oxidoreductase.
SQ SEQUENCE 463 AA; 5236 MW; CCD17293F53F812 CRC64;

Query Match 24.5%; Score 658; DB 10; Length 463;
Best Local Similarity 31.8%; Pred. No. 3.3e-39;
Matches 163; Conservative 95; Mismatches 184; Indels 70; Gaps 15;

QY 14 LLPSLLSLFLILK-----RRNRKTRFNLPPGSGMPFLGETIGYIKPYATATLGDPM 68
Db 6 LFTLLSAALEFLCLIRTAGYRRSSSTKLPLPCTMGYPYGER---FQLXSODP-NVFF 61
QY 69 QOHVSKYGTIRSNLFGPTIVSADAGLNPRIILONEGRLEFCSTPSSIGILGMSMLVL 128
Db 62 AAKQRRTGYSYKTHVLCPCVMISSPPAAKFLVLTYSKHLKPTFPAPSKRMLKROAIFEH 121
QY 129 VGDHHRMRISLNLFLSHARLRLTLKDVBRHTLFLVDSWQONSIFSADDEAKFTFNLM 188

```

```

Db 122 QGDHSLKRLKYLKAFMPDAIRN-NVPHISIAQESLNSMDGQLWTFQ-EAKTYTFNVA 179
QY 189 AKHIMSMDPGEET---EOLKKEYVFMKGVSAPLNLEGTAYHKAQSRATLKEIRK 245
Db 180 LITL-----GKDEYTRFDLRCYLLIEGYNMPTNLPGLTFHKMKARKKELAQILANI 235
QY 246 MEERKLDIKREDDDEEEVKTEDAEAMSKSDHVAKQRTDDLLGWYK-HSNLSQIIDL 304
Db 236 LSKRR-----QNPSSHT-----DLGSFWEKAGLTDQIADN 268
QY 305 ILSLFGHTSSVAIALAIFPLQACPAAVEELREEHLEIARAKKELGSELNMDYK 364
Db 269 IIGVIFRADPTASVLTWILKIYLAENPNVLEAITEOMA-I-RKDEKES-LTWGDTKKM 326
QY 365 DFTQCYINETLRGNVRELHRRALDVKYKGYDIPSGMKVLPVISAHLNDSRYDQPLN 424
Db 327 PLTYRVIQETLRATILSTFTRAEVDEYEGYLLIPKGVLPFRNTHSHSDITSNPKETDPS 386
QY 425 FNPWRMOOONNGASSSGSFSFTGNNYMPFGGPRLCAGSELAKLEMAVFTIHLVLFKN 484
Db 387 FDSRFE-----VAPKPNTEFMPFGNGTHSCPGNELAKLEISVLIHLLTKYR 433
QY 485 WEIABDDQ-----PPAFPVDPNGLPIRVSR 511
Db 434 WSIIVPDSGDIQYGFALP-----QNGLPYALAR 461

```

Search completed: March 30, 2003, 12:06:27
 Job time : 66 secs

•
•
•
•

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:06:10 ; Search time 10842.1 Seconds

(without alignments)
2955.362 Million cell updates/sec

Title: US-09-502-426a-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatcacaataatataatataat.....gagagaggaactagctcc 1101

Scoring table: IDENTITY_NDC

Gapop 10.0 ; Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank1:
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_un:*
28: em_vl:*
29: em_hlg_hum:*
30: em_hlg_inv:*
31: em_hlg_mus:*
32: em_hlg_pin:*
33: em_hlg_rtd:*
34: em_hlg_mam:*
35: em_hlg_vrt:*
36: em_hlg_vrt:*
37: em_hlg_vrt:*
38: em_hlg_vrt:*
39: em_hlg_hum:*
40: em_hlg_mus:*
41: em_hlg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1101	100.0	4818	AF044216	AF044216 Arabidops
2	1090	99.0	84196	AT13A5	AT132979 Arabidops
3	62.6	5.7	4629	AF273674	AF273674 Arabidops
4	61.4	5.6	12029	AE001372	AE001372 Arabidops
5	60.8	5.5	12330	AC117079	AC117079 Arabidops
6	60.6	5.5	3364	TSP418778	TSP418778 Arabidops
7	60.6	5.5	17317	AC020941	AC020941 Arabidops
8	60	5.4	23269	AC094291	AC094291 Arabidops
9	59.4	5.4	224635	AL733314	AL733314 Arabidops
10	59	5.4	253305	PFMAL3P7	PFMAL3P7 Arabidops
11	58.6	5.3	165669	AC114238	AC114238 Arabidops
12	58.6	5.3	192389	AC007182	AC007182 Arabidops
13	58	5.3	115489	AC117072	AC117072 Arabidops
14	57.8	5.2	27291	AC115575	AC115575 Arabidops
15	57.8	5.2	169546	AC004157	AC004157 Arabidops
16	57.6	5.2	35077	CEY38H84	CEY38H84 Arabidops
17	57.6	5.2	144784	AC009224	AC009224 Arabidops
18	57.6	5.2	188993	AC109636	AC109636 Arabidops
19	57.2	5.2	18701	AC116985	AC116985 Arabidops
20	57.2	5.2	18701	HSJ333B15	HSJ333B15 Arabidops
21	57.2	5.2	73666	CEY51H1A	CEY51H1A Arabidops
22	57	5.2	43666	AC005504	AC005504 Arabidops
23	57	5.2	104992	AC005504	AC005504 Arabidops
24	56.8	5.2	156060	AC004153	AC004153 Arabidops
25	56.8	5.2	46811	AC116550	AC116550 Arabidops
26	56.6	5.1	160759	AC117082	AC117082 Arabidops
27	56.6	5.1	6107	AX252043	AX252043 Arabidops
28	56.6	5.1	6107	AX344427	AX344427 Arabidops
29	56.6	5.1	56099	AX348832	AX348832 Arabidops
30	56.6	5.1	67970	AC115598	AC115598 Arabidops
31	56.6	5.1	86515	AC127730	AC127730 Arabidops
32	56.6	5.1	159941	AC016220	AC016220 Arabidops
33	56.6	5.1	178642	AC023122	AC023122 Arabidops
34	56.4	5.1	178642	PFMAL3P1	PFMAL3P1 Arabidops
35	56.4	5.1	65691	AC130756	AC130756 Arabidops
36	56.4	5.1	99003	PFMAL13P2_2	PFMAL13P2_2 Arabidops
37	56.4	5.1	110000	CNS01DSJ7	CNS01DSJ7 Arabidops
38	56.4	5.1	124041	AC087071	AC087071 Arabidops
39	56.4	5.1	183353	AC073320	AC073320 Arabidops
40	56.2	5.1	185296	AC115613	AC115613 Arabidops
41	56.2	5.1	163443	AC006280	AC006280 Arabidops
42	56.2	5.1	196149	AC004709	AC004709 Arabidops
43	56	5.1	204439	AF000807	AF000807 Arabidops
44	55.8	5.1	2166	AF200327	AF200327 Arabidops
45	55.8	5.1	14867	AE001398	AE001398 Arabidops

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AF044216	AF044216	Arabidopsis thaliana steroid 22-alpha-hydroxylase (DMF4) gene,	4818 bp	DNA	linear	PLAN 25-JUN-2001	Arabidopsis thaliana.	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.	Choe,S., Dilkes,B.P., Fujioke,S., Takatsuto,S., Sakurai,A. and
AF044216	AF044216	Complete cds.					Arabidopsis thaliana.		
AF044216	AF044216	GI:2935341					Arabidopsis thaliana.		

TITLE	Feldmann, K. A. The DWF4 gene of <i>Arabidopsis</i> encodes a cytochrome P450 that mediates multiple 22alpha-hydroxylation steps in brassinosteroid biosynthesis
JOURNAL	Plant Cell 10 (2), 231-243 (1998)
MEDLINE	98158690
PUBMED	9490746
REFERENCE	2 (bases 1 to 4818)
AUTHORS	Choe, S., Dilkes, P., Azpilroz, R. and Feldmann, K. A.
TITLE	Direct Submission
JOURNAL	Submitted (22-Jan-1998) Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
FEATURES	Location/Qualifiers
source	1. .4818

gene

TATA_signal

mRNA

CDS

```

/note="member of the cytochrome P450 superfamily; CYP90B1"
/codon_start=1
/product="steroid 22-alpha-hydroxylase"
/protein_id="P4C05093.1"
/db_xref="GI:2935342"
/translation="MFETEHHTLLPLLLPSLSLLFLILKRRNRTRNLPGRKSGWPIRGETIGLYIKRYPTATTLTGDPWQGVSKGYRRNLRGEPTIYASDADLNRFLIKNEGRLEPCSTPRISIGTILGKWSMLVAGDMHRRKRSISLNLSTARLRTLLIKOVERYATPLVLDWMQONSISTAQDEAKKFTFNMAKIMSDEEETDLKKEYFVMGVYASAPLNLPGTAHYAKLQSRATILTKTEFERMEERKLDIKEDDEEVEKTEDEEMKSDHDLVRKORTDDDLGWLTKHNSNSTEOIILILSLFEGHSTVAIALAIFLOACPKAVYELREHLEIARAKELGEBELNMWDKKKDLFQCVINETIRLGNVAFRLKRLKLDLRYGYDILPSGMKVLPIVISAHLNDSRTDPLFNPWMQOONNASSGSGSFSTGWCNYMPEGGGPRLCAGSELAKLEMAVFIIHLVLKFWMLAEDDPAFPFVDFNGLPIINVSRLL"

```

BASE COUNT	1566 a	727 c	888 g	1637 t
ORIGIN				

Query Match	100.0%;	Score 1101;	DB 8;	Length 4818;
Best Local Similarity	100.0%;	Pred. No. 4,3e-180;		
Matches 1101; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	AATGTACAAATTTATTAATTTATTTAGTCATAATACATGATGAAAGTCCAAAAAATTT	60
Db	32	AATGTACAAATTTATTAATTTATTTAGTCATAATACATGATGAAAGTCCAAAAAATTT	91
QY	61	TGTTAACAGAAACTTCCAAATTTTTTTTTTTATGAGAACAGAAATPACAGATGAAAC	120
Db	92	TGTTAACAGAAACTTCCAAATTTTTTTTTTTATGAGAACAGAAATPACAGATGAAAC	151
QY	121	TATTTGTTGTGCAATGCAAGTAGTATATATACATTAAAGCAAAATTTTAAAAATATATTA	180
Db	152	TATTTGTTGTGGAATGAGAGTAGTATATATACATTAAAGCAAAATTTTAAAAATATATTA	211
QY	181	GCCATATACGGCTCAAGTAGTATATACATAGGTGTAATTAATTAATGCAATGGCGCAT	240
Db	212	GCCATATACGGCTCAAGTAGTATATACATAGGTGTAATTAATTAATGCAATGGCGCAT	271
QY	241	CAGATTTGGGACCAACAATGAAACGGAATTTAAATATTTAACTTTAAAAATAAAAAT	300

Dp	272	CAGATTGGGACAAACAATGAAAAAGGAATTAAATAATTAACTTTAAATAATAAAAAATT	331
Qy	301	TGACGTAATGTGTCTTTCGACATTTAGAGGGCAAAAAAAAAAGCAATGCCAAAAGCTACG	360
Dp	332	TGAGTAATAATGTGTTTTCGACATTTAGAGGGGCAAAAAAAAAAGCAATGGCAAAAGCTACG	391
Qy	361	GGTTGACGTGCACAGTTGGGTAAATATGTAATAACTGTGCTTTAGACGACGGCTGAGT	420
Dp	392	GGTTGACGTGCACAGTTGGGTAAATATGTAATAACTGTGCTTTAGACGACGGCTGAGT	451
Qy	421	AGGGTCTCTTGACATTTTCACGTGTTACCCCTACGTCGTGAGCCACCCCTTTCCAT	480
Dp	452	AGGGTCTCTTGACATTTTCACGTGTTACCCCTACGTCGTGAGCCACCCCTTTCCAT	511
Qy	481	ATCCGAAGGGTAATTTTGGAAAAATCCCATTTTAAACCGGTTGAGACCCGACCGGATTCCT	540
Dp	512	ATCCGAAGGGTAATTTTGGAAAAATCCCATTTTAAACCGGTTGAGACCCGACCGGATTCCT	571
Qy	541	GGGATTCGTGCGAGACATTTATCAAAAATTTATAGCAGAAATGGGTTATTATTTAAAA	600
Dp	572	GGGATTCGTGCGAGACATTTATCAAAAATTTATAGCAGAAATGGGTTATTATTTAAAA	631
Qy	601	ACTCAACCTTGATCAGATTAATAAATTTCATPAACACTTTTACGATGATTCGTAGATCTA	660
Dp	632	ACTCAACCTTGATCAGATTAATAAATTTCATPAACACTTTTACGATGATTCGTAGATCTA	691
Qy	661	TCTATGACTTTTTTTTTTCTACGCGGTGGATGAAGTAATAGTACTATTAGCCAGAGA	720
Dp	692	TCTATGACTTTTTTTTTTCTACGCGGTGGATGAAGTAATAGTACTATTAGCCAGAGA	751
Qy	721	CAATTGATTATGATATATATCAATTAATCATGATATTTATGATATATAATAGCTGTTAAC	780
Dp	752	CAATTGATTATGATATATATCAATTAATCATGATATTTATGATATATAATAGCTGTTAAC	811
Qy	781	TATTTACACATCGGAGCTTCCTGCAACTTTGTTTATTAAGTAAATTAATAATAA	840
Dp	812	TATTTACACATCGGAGCTTCCTGCAACTTTGTTTATTAAGTAAATTAATAATAA	871
Qy	841	AGTATTTAAAGGAGCATTAACGACGCAACAAAGTAATGAACGAGGAACAAAGCCAT	900
Dp	872	AGTATTTAAAGGAGCATTAACGAGGCAACAAAGTAATGAACGAGGAACAAAGCCAT	931
Qy	901	GAACTCTTGTGATTAGTTAAAGCTTAATAAGAGATTTTATTAATTTTAATGACGATGA	960
Dp	932	GAACTCTTGTGATTAGTTAAAGCTTAATAAGATTTTATTAATTTTAATGACGATGA	991
Qy	961	TAAACATATATTTTCTGACTCTTTAAACCCCTCTTACAAACAGAAAGCCCTTTT	1020
Dp	992	TAAACATATATTTTCTGACTCTTTTAAACCCCTCTTACAAACAGAAAGCCCTTTT	1051
Qy	1021	CAGTAGAAGTCGATTCGCCATCTTAAAGACAAAGCCATTGAAAGAGAAAGTGAAGTAC	1080
Dp	1052	CAGTAGAAGTCGATTCGCCATCTTAAAGACAAAGCCATTGAAAGAGAAAGTGAAGTAC	1111
Qy	1081	AGAGAGAGAAACTAGCTCC	1101
Dp	1112	AGAGAGAGAAACTAGCTCC	1132

RESULT 2	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
ATT3A5/C	ATT3A5	Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.	AL132879	AL132879.2	GI:6782244	Arabidopsis thaliana.		
						Arabidopsis thaliana.		
						Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 84196)		
							Bloecher, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetler, F. and	


```

exon      /number=6
           complement(28374. .28698)
Intron    /number=7
           complement(28699. .28778)
           /number=7
           complement(28779. .28999)
exon      /number=8
           /number=8
           /gene="T3A5.50"
           complement(join(38307. .38796,38877. .39090,39189. .39270,
           40180. .40242,40337. .40401,40489. .40564,40655. .40734,
           40805. .40919,41225. .41323))
           /note="Contains Eukaryotic putative RNA-binding region
           RNP-1 signature AAI/9-186"
           /codon_start=1
           /product="U1 snRNP 70K protein"
           /protein_id="CAB62436.1"
           /db_xref="gi:5561970"
           /translation="MGDSGDPEFLNPNNAVQARAKVQNRANVLOIKMGQSHPTGLIN
           NLKLFEPREPLEYKPPPEKRCPPYGMQAFVSNFAEPDPEYAPKPEVELEPSOKR
           ERIHKLLEKVEKAADLKVDNPNMGADYKTLFVSRLYESESSEKKEEFES
           YGPIKRVHLVDQLNPKGYAFIEYHMTMDKAAVQADQOKIDGRVLDVDEGRT
           VPMRPRRIGGLGTSRYGSGEEIVGEQQPGRTSQRSEPRPREPERKSRKERE
           RSRLESHQPREBRSDRPREKHHNRDQGGRRDRDRSRRDRDTRDRGDYDRDR
           GRDRTSDHDKRDKRKKERDYEGETHEGGGSKRERDAETRGPEETRGYEDDQGD
           TDRYSHRYDKMEEDFREREYKRSKRESEYVR"
exon      /number=1
           complement(38307. .38796)
Intron    /number=1
           complement(38797. .38876)
exon      /number=1
           complement(38877. .39090)
Intron    /number=2
           complement(39091. .39188)
exon      /number=2
           complement(39189. .39270)
Intron    /number=3
           complement(39271. .40179)
exon      /number=3
           complement(40180. .40242)
           /number=4
           complement(40243. .40366)
           /number=4

Query Match      99.0%; Score 1090; DB 8; Length 84196;
Best Local Similarity 99.9%; Pred. No. 1.9e-178;
Matches 1101; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

```

Db 29741 GGTTCAGCTGTCAGTTCGGTAATAATCTAATTAACCTGTCTTTGACCGCAGCTCGTGT 29682
Qy 421 AGGGGCTCTTGACATTTTACATGTTTACCCCTACCTGAGCCACCCCTTTCCAT 480
Db 29681 AGGGGCTCTTGACATTTTACATGTTTACCCCTACCTGAGCCACCCCTTTCCAT 29622
Qy 481 ATCTTAAGGGTAATTTTGGAAATTCCTCAATTTAAACCGATTTGAGACCGTACCGACTTCGT 540
Db 29621 ATCTTAAGGGTAATTTTGGAAATTCCTCAATTTAAACCGATTTGAGACCGTACCGACTTCGT 29562
Qy 541 GGGATTCCTGAGCATTTATCAAAAATTTATAGACCAATGGCTTTATTAATTTAAA 600
Db 29561 GGGATTCCTGAGCATTTATCAAAAATTTATAGACCAATGGCTTTATTAATTTAAA 29502
Qy 601 ACTCAACACTGATCAGATAAATTTCAATAACACTTTAGCATGATTCGATCA 660
Db 29501 ACTCAACACTGATCAGATAAATTTCAATAACACTTTAGCATGATTCGATCA 29442
Qy 661 TCTAATGACTTTTTTTTTTTCTAACACAGGTGATGAAAGTTATAGTACTATTAAGCCAGAGA 720
Db 29441 TCTAATGACTTTTTTTTTTTCTAACACAGGTGATGAAAGTTATAGTACTATTAAGCCAGAGA 29382
Qy 721 CAATGATATATGATATATATCCATTAATCCATGATATTTATGATATTAATAGCTTTAAAC 780
Db 29381 CAATGATATATGATATATATCCATTAATCCATGATATTTATGATATTAATAGCTTTAAAC 29322
Qy 781 TATTCACAGATCGACAGCTTTCTGCAACTTTGTTTAAATTAAGAGTTAAATTAATAA 840
Db 29321 TATTCACAGATCGACAGCTTTCTGCAACTTTGTTTAAATTAAGAGTTAAATTAATAA 29262
Qy 841 AGTATTAAGAGCATTAAGAGGCAACAAAGTAATGAACAGGAGAAACAAAGCCAT 900
Db 29261 AGTATTAAGAGCATTAAGAGGCAACAAAGTAATGAACAGGAGAAACAAAGCCAT 29202
Qy 901 GAAGCTCATTTGTTA-GTTTAAGCTTAATGAAGATTTTATTAATTTAATGACAGATG 959
Db 29201 GAAGCTCATTTGTTAAGCTTAATGAAGATTTTATTAATTTAATGACAGATG 29142
Qy 960 ATTAACAATTAATTTTCTGACTCTTTAAACCCCTCTTAAACAAGCAAGCTCCCTTT 1019
Db 29141 ATTAACAATTAATTTTCTGACTCTTTAAACCCCTCTTAAACAAGCAAGCTCCCTTT 29082
Qy 1020 TCAGTGAAGTCCGATTCCTTAAGACAAAGCCATTAAGAAAGAAAGTGAAGA 1079
Db 29081 TCAGTGAAGTCCGATTCCTTAAGACAAAGCCATTAAGAAAGAAAGTGAAGA 29022
Qy 1080 GAGAGAGAGAACTAGCTCC 1101
Db 29021 GAGAGAGAGAACTAGCTCC 29000

RESULT 3
AF273674
LOCUS AF273674 4629 bp DNA linear INV 24-APR-2001
DEFINITION Plasmodium falciparum DNA-dependent RNA polymerase (TENA) gene,
complete cds; nuclear gene for probable mitochondrial product.
ACCESSION AF273674
VERSION AF273674.1 GI:9857984
KEYWORDS
SOURCE
ORGANISM Plasmodium falciparum.
Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
AUTHORS Li,J., Muga,J.A., Cermakian,N., Cedergren,R. and Feagin,J.E.
TITLE Identification and characterization of a plasmodium falciparum RNA
polymerase gene with similarity to mitochondrial RNA polymerases
Mol. Biochem. Parasitol. 113 (2), 261-269 (2001)
JOURNAL MEDIANL 11295180
PUBMED 21192559
REFERENCE 2 (bases 1 to 4629)
AUTHORS Li,J., Muga,J.A., Cermakian,N., Cedergren,R. and Feagin,J.E.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2000) Seattle Biomedical Research Institute, 4

```


BASE COUNT	ORIGIN
4945 a	1191 c 943 g 4950 t
NDRKSNQOANSODCEPFLFGLKSLDRKTVTEEDLEPMILTEPGKLERKDYDILNMM YSHLDRKFNKMSLSLIGCKLAHEKRLPFRKMKMKWCCGHVDELLIKHDDY NSJNYINNNESSREDFLFLNMKIKHSHWTFITPMFTICKKISLENNRNYTN"	

	Db	18550	AAAAAAAAACAATAAAAAAAAAAAAAAAAAAAATTAATACTGGAATCGAATTACT	18491
Oy	723	ATT-GATTATAGATTTATCCATTATTCATGATATTTTGTATATAAAGCTTTAAACT	781	
Db	18490	CTTCATATATATGATATCACTACTATAAAAGTTTTTAATATATTTATGTTAACTCCT	18431	
Oy	782	ATTTCAGATGCACCTTCTGCAACTTTTGTTTTTAATTTAAGAGTTTAAATAAATA	841	
Db	18430	CTTTAAAAAACAAAATTAAGAAAATTAACCAAAAAAAAAAAAAAAAAAAAAAAA	18371	
Oy	842	GTAATTAAAGNAGCATTAACGAGGCAACAAAAGTATGAACGGAACCAACCATG	901	
Db	18370	AATAAAACCCATTAATAATTTTAAAGAATAAATATATGATATTTTAATAATAAT	18311	
Oy	902	AAGCATTTGGTAGTTTAAGCTTAATTAAGAAGATTTTATTAATTTT	949	
Db	18310	TTAATTTTATTTTATTTTATTTTGTATTTTATTTATTTATTTATTTT	18263	
RESULT 6				
LOCUS	TSP418778/c	3364 bp	RNA	linear VRL 02-MAY-2002D
DEFINITION	Tomato spotted wilt virus NSs gene and N gene, genomic RNA, isolate			
ACCESSION	U418778			
VERSION	U418778.1	GI:20429081		
KEYWORDS	N gene; N protein; NSs gene; NSs protein.			
SOURCE	Tomato spotted wilt virus.			
ORGANISM	Tomato spotted wilt virus			
REFERENCE	1			
AUTHORS	Heinze, C., Willingmann, P. and Adam, G.			
TITLE	Short intergenic regions of the S RNAs of Tomato spotted wilt tospovirus - not a species characteristic			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 3364)			
AUTHORS	Adam, G.			
TITLE	Direct Submission			
JOURNAL	Submitted (05-NOV-2001) Adam G., Pflanzenenschutz, Institut fuer			
REFERENCE	Angewandte Botanik, Ohnhorststrasse 18, 22609 Hamburg, GERMANY			
FEATURES	Location/Oualifiers			
source	1..3364			
	/organism="Tomato spotted wilt virus"			
	/viation			
	/isolate="10HK96"			
	/specific_host="Virginia tobacco"			
	/db_xref="taxon:11613"			
	/country="Bulgaria"			
	89..1492			
gene	/gene="NSs"			
CDS	89..1492			
	/gene="NSs"			
	/codon_start=1			
	/product="NSs protein"			
	/protein_id="CAD11445.1"			
	/db_xref="GI:20429082"			
	/translation="MSSSVYESIIQTFRASVWGSTASKAVVDYSWIHELGTSLVYOT QLSDSRKSGSFYTAKVGNLPCEEELLSQHYIPIFDIDFISINDSYLALSCVS NTWNGEHTKHQHGLKVLSPOLHISGSTMRSDKDPEOERDIINDRIEANGSG SLGSEHTYKIETFCYNOMALGXVNLSPNNVHEMLYSPFKPMOVESNRNTVASLV RSLLTSLENNIMRNSQAFYKASIDSHEKLSIMLBRYVKLOYSIQKLFEKAGDETNTK FYSLIVALETPHNSTFALNTSVICKHOLPIRKCAPSELNVESDLKEPNIIVHPSTY PORVALLETTHSTFAVOLCNNDQEDVIITLNNYEELTPKLDIGERTLANSIEDICKRR KYFLSKTECLPSPTOTMSYLDSIQIPSMRIDFARGIKISIPGVSAKSILKIDLGS IKKKSGKSIDPVASGSK"			
	complement(2436..33212)			
gene	/gene="N"			
CDS	complement(2436..33212)			
	/gene="N"			
	/function="structural protein"			
	/codon_start=1			
	/product="N protein"			
	/protein_id="CAD11446.1"			

BASE COUNT	1002	a	563	c	489	g	1310	t	
ORIGIN									
Query Match	5.5%	Score	60.6;	DB	14;	Length	3364;		
Best Local Similarity	48.2%;	Pred. No.	0.38;						
Matches	171;	Conservative	0;	Mismatches	184;	Indels	0;	Gaps	0;
QY	1	AATCAACAATTATTAATTAGTCATTAACATGCTACGAAAGTTCCAAAAAATTT	60						
Db	2100	AATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAAT	2041						
QY	61	TGTTATACGAAGCTCCAAATTTTTTTTTTTTATGGAACAAGAAATACAGATGAAAC	120						
Db	2040	AAAAAATAAAAAGTAAAAAATAGCAATAAAAAATAAAAAATAAAAAATAAAAA	1981						
QY	121	TATTTTGTGTGATGAGAGTAGTAAATATACATTAAAGCAATTTTAAAAATATATA	180						
Db	1980	AATAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAA	1921						
QY	181	GCCTATACGCCCTCAAGATGTTATCTAGTAGGTGTAATTAAATATGATGTCGATT	240						
Db	1920	AAATAAAAAATAAAAAATAAAAAATGAAAAATAAAAAATAAAAAATAAAAA	1861						
QY	241	CAGATTGGAGCAACAATGAAAGCGATTAAATATTAATTAATTAATTAATTA	300						
Db	1860	TAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAA	1801						
QY	301	TGAGTAAATGTGTTTCTGACTATTGAGGGCCAAAAAAGACAAATGCCAAAGT	355						
Db	1800	ATAAAAATAAAAAATAAAAACAAAAAACAACAAAAACAAACCCAATT	1746						
RESULT 7									
AC020941/c	171317	bp	DNA	linear	PRI	01-JUN-2001			
LOCUS									
DEFINITION	Homo sapiens chromosome 5 clone CTD-231219, complete sequence.								
ACCESSION	AC020941								
VERSION	AC020941.5 GI:14277273								
KEYWORDS	HTG.								
SOURCE	Homo sapiens.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.								
AUTHORS	1 (bases 1 to 171317)								
JOURNAL	DOE Joint Genome Institute and Stanford Human Genome Center.								
REFERENCE	Unpublished								
AUTHORS	2 (bases 1 to 171317)								
JOURNAL	DOE Joint Genome Institute.								
TITLE	Direct Submission								
REFERENCE	Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA								
AUTHORS	3 (bases 1 to 171317)								
JOURNAL	DOE Joint Genome Institute and Stanford Human Genome Center.								
TITLE	Direct Submission								
REFERENCE	Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA								
AUTHORS	On Jun 1, 2001 this sequence version replaced gi:12830142.								
JOURNAL	Draft Sequence Produced by DOE Joint Genome Institute								
COMMENT	www.jgi.doe.gov								
FEATURES	Finishing Completed at Stanford Human Genome Center								
SOURCE	www.sngc.stanford.edu								
	Quality: Phrap Quality >=40 99.7% of Sequence:								
	Estimated Total Number of Errors is 2.7.								
	Location/Qualifiers								
	1..171317								
	/organism="Homo sapiens"								

* 10040 10139: gap of unknown length
* 10140 11793: contig of 1654 bp in length
* 11794 11893: gap of unknown length
* 11894 13371: contig of 1478 bp in length
* 13372 13471: gap of unknown length
* 13472 14864: contig of 1393 bp in length
* 14865 14965: gap of unknown length
* 14966 16460: contig of 1496 bp in length
* 16461 16560: gap of unknown length
* 16561 17663: contig of 1103 bp in length
* 17664 19281: contig of 1518 bp in length
* 19282 19381: gap of unknown length
* 19382 21112: contig of 1731 bp in length
* 21113 21112: gap of unknown length
* 21213 22833: contig of 1621 bp in length
* 22834 24214: contig of 1281 bp in length
* 24215 24314: gap of unknown length
* 24315 25527: contig of 1213 bp in length
* 25528 25628: gap of unknown length
* 25629 27220: contig of 1593 bp in length
* 27221 27320: gap of unknown length
* 27321 28926: contig of 1606 bp in length
* 28927 29026: gap of unknown length
* 29027 30363: contig of 1337 bp in length
* 30364 30463: gap of unknown length
* 30464 31665: contig of 1202 bp in length
* 31666 31765: gap of unknown length
* 31766 33349: contig of 1584 bp in length
* 33350 33449: gap of unknown length
* 33450 36317: contig of 2868 bp in length
* 36318 36417: gap of unknown length
* 36418 37778: contig of 1361 bp in length
* 37779 37878: gap of unknown length
* 37879 39851: contig of 1973 bp in length
* 39852 39852: gap of unknown length
* 39853 42802: contig of 2851 bp in length
* 42803 42902: gap of unknown length
* 42903 45682: contig of 2780 bp in length
* 45683 45782: gap of unknown length
* 45783 49268: contig of 3486 bp in length
* 49269 49368: gap of unknown length
* 49369 51428: contig of 2060 bp in length
* 51429 51528: gap of unknown length
* 51529 54571: contig of 3043 bp in length
* 54572 54671: gap of unknown length
* 54672 57546: contig of 2875 bp in length
* 57547 57646: gap of unknown length
* 57647 60741: contig of 3094 bp in length
* 60741 60840: gap of unknown length
* 60841 63916: contig of 3076 bp in length
* 63917 64016: gap of unknown length
* 64017 67000: contig of 2984 bp in length
* 67001 67100: gap of unknown length
* 67101 70219: contig of 3119 bp in length
* 70220 70319: gap of unknown length
* 70320 72950: contig of 2631 bp in length
* 72951 73050: gap of unknown length
* 73051 76802: contig of 3752 bp in length
* 76803 76902: gap of unknown length
* 76903 79990: contig of 3088 bp in length
* 79991 80090: gap of unknown length
* 80091 82745: contig of 2655 bp in length
* 82746 82845: gap of unknown length
* 82846 87725: contig of 4880 bp in length
* 87726 87825: gap of unknown length
* 87826 91178: contig of 3353 bp in length
* 91179 91278: gap of unknown length
* 91279 94533: contig of 3255 bp in length
* 94534 94634: gap of unknown length
* 94635 99871: contig of 5238 bp in length
* 99872 gap of unknown length

* 99972 103742: contig of 3771 bp in length
* 103743 103842: gap of unknown length
* 103843 108877: contig of 5035 bp in length
* 108878 108977: gap of unknown length
* 108978 114019: contig of 5042 bp in length
* 114020 114119: gap of unknown length
* 114120 119231: contig of 5112 bp in length
* 119232 119331: gap of unknown length
* 119332 125649: contig of 6318 bp in length
* 125650 125750: gap of unknown length
* 125751 132687: contig of 6938 bp in length
* 132688 132788: gap of unknown length
* 132789 138741: contig of 6954 bp in length
* 138742 139841: gap of unknown length
* 139842 146349: contig of 6508 bp in length
* 146350 155325: contig of 8876 bp in length
* 155326 155425: gap of unknown length

Query Match 5.4%; Score 60; DB 2; Length 233269;
Best Local Similarity 51.3%; Pred. No. 0.2;
Matches 138; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

OY 6 ACAATTTTATATTTGTCATATACATGCAAGTTCACAAAATTTTGTGTA 65
Db 1391 AAAAAATTTTGTGTTTTTTTAAAGTTAAAAATTTTAAAAATTTTGT 1450
OY 66 ACAGAACTCCAAATTTTTTTTTTTTAVGCAACAGAAATACAGTAAACTATT 125
Db 1451 TTTTAAAAAAATTTTGTGTTTTTTTGAAGAAAAATTTTAAAAATTTATTT 1510
OY 126 TGTGTGGAATGAGAGTATATATCATTAAGCAATTTTAAAAATTTATTAAGCCTA 185
Db 1511 TTTTAAAAAAATTTTGTGTTTTTTTGTGCAAAAAATTTTAAAAATTTT 1570
OY 186 TACGGCTCAAGTGTATCTAGTATGTAATTAATGCTGTGTCATTCAGAA 245
Db 1571 TTTTAAAAAAATTTTGTGTTTTTTTGTGCAAAAAATTTTAAAAATTTT 1630
OY 246 TTGGGACAAACATGAAACGCAATTTAA 274
Db 1631 TTTTAAAAAAATTTTAAAAATTTTAAAA 1659

RESULT 9
AL732314/c 224635 bp DNA linear HTG 17-AUG-2002
LOCUS
DEFINITION Homo sapiens chromosome X clone RP13-465B17, *** SEQUENCING IN
ACCESSION AL732314 GI:22415930
VERSION AL732314.10
KEYWORDS HTG; HTGS_PHRASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 224635)
AUTHORS Whitehead,S.
TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 21, 2002 this sequence version replaced gi:22204483.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: B4465B17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads

11031	1369	contig of 2439	bp in length
13470	13569	gap of unknown length	
13570	14675	contig of 1106	bp in length
14676	14775	gap of unknown length	
16103	16102	contig of 1327	bp in length
16203	16202	gap of unknown length	
18321	18320	contig of 2118	bp in length
18431	18420	gap of unknown length	
20659	20658	contig of 2238	bp in length
20759	20758	gap of unknown length	
22074	22073	contig of 1315	bp in length
22174	22173	gap of unknown length	
24784	24783	contig of 2610	bp in length
24884	24883	gap of unknown length	
27461	27460	contig of 2577	bp in length
27461	27460	gap of unknown length	
29144	29143	contig of 1583	bp in length
29244	29243	gap of unknown length	
31382	31381	contig of 2138	bp in length
31482	31481	gap of unknown length	
33732	33731	contig of 2250	bp in length
33732	33831	gap of unknown length	
33832	36867	contig of 3036	bp in length
36868	36867	gap of unknown length	
36968	36968	contig of 2917	bp in length
39885	39884	gap of unknown length	
39985	42654	contig of 2670	bp in length
42655	42754	gap of unknown length	
42755	44617	contig of 1863	bp in length
44618	44717	gap of unknown length	
44718	46344	contig of 1527	bp in length
46345	46344	gap of unknown length	
46345	49733	contig of 3389	bp in length
49734	49833	gap of unknown length	
49834	52093	contig of 2260	bp in length
52094	52193	gap of unknown length	
52194	54926	contig of 2733	bp in length
54927	55026	gap of unknown length	
55027	55824	contig of 3398	bp in length
58455	58524	gap of unknown length	
61514	61513	contig of 2989	bp in length
61514	61613	gap of unknown length	
63685	63685	contig of 2072	bp in length
63686	63785	gap of unknown length	
63786	66780	contig of 2995	bp in length
66781	66880	gap of unknown length	
70125	70124	contig of 3244	bp in length
70224	70224	gap of unknown length	
74002	74001	contig of 3777	bp in length
74102	74101	gap of unknown length	
74102	77381	contig of 3280	bp in length
77382	77481	gap of unknown length	
77482	80569	contig of 3088	bp in length
80570	80669	gap of unknown length	
80670	83635	contig of 2966	bp in length
83636	83735	gap of unknown length	
83736	87660	contig of 3925	bp in length
87661	87760	gap of unknown length	
87761	90614	contig of 2854	bp in length
90615	90714	gap of unknown length	
90715	95460	contig of 4746	bp in length
95461	95560	gap of unknown length	
95561	100030	contig of 4470	bp in length
100031	100130	gap of unknown length	
100131	106875	contig of 6745	bp in length
106876	106975	gap of unknown length	
106976	111951	contig of 4976	bp in length
111952	112051	gap of unknown length	
112052	116127	contig of 4076	bp in length
116128	116127	gap of unknown length	
116228	120328	contig of 4101	bp in length
120329	120428	gap of unknown length	
120429	124423	contig of 3995	bp in length

FEATURES	source	Location/Qualifiers
* 124424	124523:	gap of unknown length
* 124524	132057:	contig of 7534 bp in length
* 132058	132157:	gap of unknown length
* 132158	139108:	contig of 6951 bp in length
* 139109	139208:	gap of unknown length
* 139209	149165:	contig of 9957 bp in length
* 149166	149265:	gap of unknown length
* 149266	165669:	contig of 16404 bp in length.
FEATURES	source	Location/Qualifiers
BASE COUNT	48928 a 31280 c 30162 g 48184 t 7115 others	
ORIGIN	5.3%; Score 58.6; DB 2; Length 165669; Best Local Similarity 53.8%; Pred. No. 0.38; Matches 142; Conservative 0; Mismatches 120; Indels 2; Gaps 1;	
QY	40	AGAAAGTCCAAAAATTTTGGTACAGAACTCCAAATTTTTTTTTTATGAGAAC 99
DB	1457	AAAAAATTTAAAAAATTTATTAATAAAAAAATTTTATTTTAAAAA 1516
QY	100	AAGAAATTAACAGATAGAAACATTTTGTGTGCGAATGGAAGTAAATATACATTAGC 159
DB	1517	AAAAAAAAAAAAAAAAATTTTTTTTTTGGGAAAAAATVGGGAAAAAATTTTAA 1576
QY	160	AAATTTTAAAAATTTATTAAGCTATACGCGCTCAAGATATGTATCATGTAGTGTAA 219
DB	1577	ATTATTAATTTAGAGATTAATAAAAAAATTTTAAATTTTAAAAAATTTAAAAA 1636
QY	220	TTAATAATGCATGTGCGATT--CAGAAATGGGACACAAATGAAACGGAATTTAAATAT 277
DB	1637	ATAATAATTTTGTATAATTAATTTATTAATAATAATAATAATAATAATAATAATAAT 1696
QY	278	TAACTTAAAAATTAATTAATAATTT 301
DB	1697	AAAAATTAATTTTAAATAATAATTT 1720
RESULT 12		
LOCUS	AC007182	192389 bp DNA linear PRI 25-AUG-1999
DEFINITION	Homo sapiens chromosome 14 clone BAC 507E23 map 14q24.3, complete sequence.	
ACCESSION	AC007182	
VERSION	AC007182.3	GI:5708446
KEYWORDS	HTG.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 192389)	
TITLE	Rosen, L., Madan, A., Qin, S., Abbasi, N., Bardaran, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S., Madan, A., Ratcliffe, A., Shaffer, T. and Hood, L.	
JOURNAL	Sequencing of human chromosome 14	
REFERENCE	Unpublished	
AUTHORS	2 (bases 1 to 192389)	
TITLE	Young, J., Rosen, L., Madan, A., Qin, S., Abbasi, N., Dors, M., Dickhoff, R., James, R., Loretz, C., Lasky, S., Madan, A., Prescott, S., Ratcliffe, A., Shaffer, T. and Hood, L.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (30-MAR-1999) Multimegabase Sequencing Center, University of Washington, PO BOX 357730, Seattle, WA 98195, USA	
AUTHORS	3 (bases 1 to 192389)	
TITLE	Rosen, L., Madan, A., Qin, S., Abbasi, N., Bardaran, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S., Madan, A., Ratcliffe, A., Shaffer, T. and Hood, L.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (25-AUG-1999) Multimegabase Sequencing Center, University of Washington, PO BOX 357730, Seattle, WA 98195, USA	

/note="ORF_ID:dd_01651"
/codon_start=1
/product="Sequence 3 from Patent WO0075321. 6/101"
/protein_id="AA03154.1"
/db_xref="GI:20976565"
/translation="MNQAAVESINSLKSTADYPPKGYHTELESDSYIDHEHFSNG
ATFHNPAVAAYLEONLAFHESGSAITSGALVTRGVGRSPKPKRIYKEPSQDDI
WMPVPAIMDDLSFMIINREAIIDYINTQEKIYIDYAGMDPKYRIKYAICARAYHA
LGMHMLIRPANRELRNRPEDDYIYNAGCPANRNGDSSSSSIDPARKEMVI
LGTQYAGEKKGILITIMYIMKMGVLPFHSSCNQANRNGDSSSSSIDPARKEMVI
DINRELIGDEHEDTDCGFENEGGCAKCIDLSREKEPEIFDAIKFSAVLENNYNE
YSRKYDNDVSTENTRCAPLEHHPNPKFPAIACHPKNIMLTCDAAGCILLPVSRLD
ANQVAHPLQGTAKVAGTEVGVTPPTAFSSCGEPPIVMPHPIYKMLASOLKHS
AAAMILNMGWGSGSGVSGRIATRAIIDAHSGLKLEKIPTEKMDYGFQVPSCP
GVSEILMPINWADKERYSTMRLATLFTENEFQDKASPELVAAGPILPQ"
join(28969..29084,29176..29317,29453..30610)
/note="ORF_ID:dd_00794"
/codon_start=1

Query Match 5.3%; Score 58; DB 2; Length 115489;
Best Local Similarity 52.0%; Pred. No. 0.52;
Matches 155; Conservative 0; Mismatches 140; Indels 3; Gaps 1;

QY 1 AATACAAATTTAATATTAGTCAATACATGATAGAAAGTCCAAAAAATTT 60
DB 66587 AAAAAAATAATATATATATATATATATATATATATATATATATATAT 60
QY 61 TGTAAAGAACTTCAATTTTATTTTATGAAACAGAAATAACAGATAGAAAC 120
DB 66527 AAGAAAAAATTCATCAATGATGTTTAAAGCAAAAAAATAAAT 66471
QY 121 TATTTTGTGAGAGAGAGATGATATATATATATATATATATATATATAT 180
DB 66470 TTTATTTATTTGATATATATATATATATATATATATATATATATATAT 66411
QY 181 GCGATATAGCGCTCAATATGTTATGATGTTATATATATATATATATATAT 240
DB 66410 TAGGAACACATCAATTTTATTTTATTTGCGTCCACATGATATATATAT 66351
QY 241 CAGATTTGGAGACAAATGAAAGGATTTAATATATATATATATATATATAT 298
DB 66350 TCCACTATTTACCAATTTTAAAAATATATATATATATATATATATATAT 66293

RESULT 14
AC11575/c 27291 bp DNA linear HTG 21-MAR-2002
LOCUS dictyostelium discoidium chromosome 2 map 1180800-1208089 strain
DEFINITION AX4 *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AC11575
VERSION AC11575.1 GI:19569867
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Dictyostelium discoidium.
ORGANISM Dictyostelium discoidium.
Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.

REFERENCE 1 (bases 1 to 27291)
AUTHORS Gloeckner,G., Eichinger,L., Szafinski,K., Pachebat,J., Dear,P.,
Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guiso,R., Kumpf,K.,
Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
Noegel,A.A.
TITLE Sequence and Analysis of Chromosome 2 of Dictyostelium
JOURNAL The Dictyostelium Genome Sequencing Consortium
AUTHORS Baumgart,C.
REMARK Direct Submission
COMMENT Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
CDS predictions from Genedit may contain errors. Further information
is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shml)

Funding
Agency : Deutsche Forschungsgemeinschaft. (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
the accession number will be preserved.
Location/Qualifiers
1..27291
/organism="Dictyostelium discoidium"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="1180800-1208089"
complement(join(163..380,1237..2062,2157..2468))
/note="ORF_ID:dd_00011"
/codon_start=1
/product="Hypothetical protein yfew precursor"
/protein_id="AA192187.1"
/db_xref="GI:19569868"
/translation="MKYLNLILFEFVILICVYSLKSPKSKIDVPRLNQOVRPSD
ETWSEVDTELEAILENGTTPGCVAVLNGGVLYASAKGSFTYIGTPIINDEVPVPE
LSTLFDAASCSTYACTTSAOYORSELMLDSDITDDELGEVNGKSGISIRKCLL
HNSGFYDPPMPFNATFENCPAEEYEPPEPEFCOSOIYOSILNOSLAFIGSTIYVS
DLMNMTLMYVGNLAKOENYITADQLTRGCDKAPENAOCTEAYVRYVFNALNLY
TNLIPQVYLAPOCAPENDYVYHHTITIGVSDGNAYALGGSAGVSVNVEEMIF
MYSLMREANSOYLNSTTVQYFTFETYNHOSRALQMNNDSPDEGGLACGTLSA
KTMHGLITETMLCGDERELIVILLTRNYRPPSPNIKENYKPFSTLYQOYYSII
N"

complement(join(2838..3347,3546..3650))
/note="ORF_ID:dd_00013"
/codon_start=1
/product="HSPC305 (FRAGMENT)". 10/100"
/protein_id="AA192188.1"
/db_xref="GI:19569869"
/translation="MTDNNIFGTNRKFNDSIELINKLETPROKILRYNERIPPESE
ENLQNLKFTLSLEFGKILECCSFIEQGTAYSLSPNNLYNOLKTKMLDNDTSCPOS
VWEDNSEHNLNFTQSIAPLOLNEIGLHOMSSPAIKRNASAIMELNANNNNNN
NN
join(3837..3886,4358..5195)
/note="ORF_ID:dd_00015"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA192189.1"
/db_xref="GI:19569870"
/translation="MGIFRSLFTIIGIILNCSGKICLNQDSCNOSYIEKINKTKO
QOOLLNKISINKLITNENKSPKSKIKHFFINKESNNIIVYSIVKNNINGNDILNRY
KOLPKNKIEITDKYITLNNSNFPCLNKNNGKPCYOLISTLVYNNNNYNNYLSKI
IYPRDENVGKFDNNMLNKKSVYIQQHSLKIALGTINNNEVWVSAAYINIDPEKCI
PKSISKNSQSILITNSRDFNRCLTNSGTCLPAFNENNTVIVPIYPTCPKGYLFFFG
NNGKILFENCDANFLDTRY"
join(6320..6707,6802..7274)
/note="ORF_ID:dd_00017"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA192190.1"
/db_xref="GI:19569871"
/translation="MIRKFCISALIFIFISIGSTLIRPBYRIMDENCIRAR
POYSAOKYNNENSTALISFTYDPGYKNSYLSFEARSLLIFGSIOPNPFTAKDL
KVRVYKSLPLGNPASATDKYIFGDNKRVVTPPCASTDVLNLRKQIVDSILML
YSKNKGYFDSAMIGSKVRSDDYGLIGGQFSSGKTTISNAFTLHPDQIGKDSLPI
LKCQGSIMYTSRKNRCLASPTCTQSSFTLGMPLCPAGYRILQDFATEANGCPKY
CDPFFVCKTH"
join(7879..8275,8373..8845)
/note="ORF_ID:dd_00018"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA192191.1"
/db_xref="GI:19569872"
/translation="MINKYLLLEFTEFTFVSSSITKANFYKYKVMENOCVY
PCKCYIOLKLNENKOKLIDLEFPGNLKTLTGDSKTLIVLGTQSSKPEPNA
TDLVTVYVYSLPGNKQISTDYKYMGGDNGGCTKSPCNVYALALNHSQDITWT
SOPKRVNGFLDSVWLSKNNIRSDPGLTGGKIKNGVYSINSFTYLPDPKCPCL
PLKCYVGNMNTYSRDNRCLVSPCKTKREVCTLSIPLCNKGYRIVSPSTELNGCPK

CDS

CDS

CDS

CDS

CDS

REFERENCE 2 (bases 1 to 169546)

AUTHORS Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT

On Aug 12, 2000 this sequence version replaced gi:8810447.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

1 23466: contig of 23466 bp in length
* 23467 23666: gap of unknown length
* 23667 169546: contig of 145880 bp in length.
Location/Qualifiers
1..169546
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"
/clone="PFVAC293"
/clone="3D7"

BASE COUNT 69871 a 15381 c 15705 g 68389 t 200 others
ORIGIN

Query Match

Best Local Similarity 49.9%; Pred. No. 0.52; Length 169546;
Matches 173; Conservative 0; Mismatches 172; Indels 2; Gaps 1;

OY 14 TTAAATTTAGTCATACATGATGAGAAAGTTCAGAAAAATTTGTTACAGAAAC 73
Db 13340 TTATGAGAAATTAAGATGATATATATATATATATATATATATATATATAC 13281
OY 74 TTCCAAATTTTATTTTATGACACAGAAATTAACAGATAGAAACTATTTGTTGCG 133
Db 13280 TTATATTTTATATGTTATATATAAAGAACTTATGTCACAACTTATATATATA 13221
OY 134 AATGAGATGATATATATATATATGCAAAATTTTAAAAATTTATATAGCC--TATAGCG 191
Db 13220 TATTA 13161
OY 192 CTCAAAGTATGTTATCTAGTGTATATATATATATATATATATATATATATGGA 251
Db 13160 CTACATATCTTTAT 13101
OY 252 CAACAAATGAAACGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 311
Db 13100 GAAAAAAGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 13041
OY 312 GTTTTCGACTATGAGGGGCAAAAAAAGCAATGCCAAAGCTA 358
Db 13040 TTTTTCGTTTATGATTAACAAATTTGTACAAAGCAAAAAACA 12994

Search completed: March 30, 2003, 07:31:25
Job time : 12061.1 secs

1

GenCore version 5.1.4-p5-A578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:05:35 ; Search time 219.917 Seconds
(without alignments)
11274.505 Million cell updates/sec

Title: US-09-502-426a-1_COPY_2102_3202

Perfect score: 1101
Sequence: 1 aatcaacaattattatattat.....gagagagaagaactagctcc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

```
1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1101	100.0	6888	21	AAA59599	DNA encoding a cyt
2	56.6	5.1	6107	24	ABL70390	Chemically treated
3	56.6	5.1	6107	24	AA561342	Human gene regulat
4	56.6	5.1	6107	24	ABK31431	Signal transductio
5	54.2	4.9	875	21	AAA01920	Human colon cancer
6	54.2	4.9	6113	24	ABJ32803	Human immune syste
7	53.8	4.9	37973	24	ABJ34197	Human immune syste
8	53.8	4.9	83391	24	ABG67094	Human anglogenesis
9	53.4	4.9	17131	24	ABJ33053	Human immune syste

10	53.2	4.8	8170	24	ABK28257	DNA transcription
11	53	4.8	54108	24	ABK22782	Human high bone ma
12	53	4.8	57273	24	ABK22784	Human high bone ma
13	53	4.8	66933	22	ABA82625	Human HBM gene reg
14	53	4.8	72049	22	ABA82623	Human HBM gene reg
15	52.6	4.8	16228	24	ABJ70459	Chemically treated
16	52.6	4.8	16228	24	AA561424	Human gene regulat
17	52.4	4.8	4501	21	ABK33968	Human DNA for stag
18	52.4	4.8	5940	21	AAA70105	Plasmodium falcipa
19	52.2	4.7	7746	24	ABK40047	Human chemically p
20	52.2	4.7	7746	24	ABL33856	Human immune syste
21	52.2	4.7	78925	21	AA899888	Human FN gene. Ho
22	52	4.7	5309	22	AA546527	Tumour suppressor
23	52	4.7	5309	22	ABK40039	Human chemically p
24	52	4.7	5309	24	ABL33736	Human immune syste
25	52	4.7	6636	24	ABN80023	Human chemically m
26	52	4.7	7459	24	ABK31382	Signal transductio
27	51.8	4.7	641	24	ABO56694	Human colon cancer
28	51.6	4.7	47108	24	ABK31511	Signal transductio
29	51.4	4.7	4041	21	AAK70170	Plasmodium falcipa
30	51.2	4.7	3738	21	AAA70178	Plasmodium falcipa
31	51.2	4.7	5690	22	AA545368	Chemically pretrea
32	51.2	4.7	5690	22	ABK28205	DNA transcription
33	51.2	4.7	23695	24	ABO66981	Human anglogenesis
34	51	4.6	10286	22	AA545309	Chemically pretrea
35	51	4.6	10286	24	ABK28148	DNA transcription
36	50.8	4.6	5880	22	AA546331	Tumour suppressor
37	50.8	4.6	5880	22	ABK28177	DNA transcription
38	50.8	4.6	6254	24	ABJ33621	Human immune syste
39	50.6	4.6	12405	22	AA545330	Chemically pretrea
40	50.6	4.6	12405	24	AA561143	Human gene regulat
41	50.6	4.6	12405	24	ABK28169	DNA transcription
42	50.4	4.6	5689	22	AA545384	Chemically pretrea
43	50.4	4.6	5689	22	AA546426	Tumour suppressor
44	50.4	4.6	5689	24	ABK28226	DNA transcription
45	50.4	4.6	8085	22	AA546479	Tumour suppressor

ALIGNMENTS

RESULT 1	AAA59599	standard; DNA; 6888 BP.
ID	AAA59599	
AC	AAA59599	
XX	14-NOV-2000	(First entry)
DT	DNA encoding a cytochrome P450 enzyme designated DMF4.	
XX	DMF4, cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;	
KW	Plant phenotype; cell elongation; ss.	
XX	Arabidopsis sp.	
OS	Key	Location/Qualifiers
XX	promoter	1..3203
FT	TATA-signal	/tag- a
FT		3060..3125
FT	CDS	/tag- b
FT		3203..6110
FT		/tag- c
FT		/product= "DMF4"
FT		/note= "contains introns"
FT	exon	3203..3423
FT		/tag- d
FT	intron	3424..3503
FT		/tag- e
FT	exon	3504..3828
FT		/tag- f
FT	intron	3829..3913
FT		/tag- g

[illegible][illegible]

RESULT 9
ABL33053/c
ID ABL33053 standard; DNA; 17131 BP.
XX

AC ABL33053;
XX

26-MAR-2002 (first entry)

XX associated gene SEQ ID NO: 1026.

KW antiarteriosclerotic; vasese; cytosine methylation; antisthmatic
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antihemmatic; antarthritic; antidyslipetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

Homo sapiens.

WO200200928-A2.

03-JAN-2002

02-JUL-2001; 2001WO-EP07537.

01-SEP-2000; 2000DE-1043825

(EPIG-) EPIGENOMICS AG.

Olex A, Plepenbrock C, Berlin K;

WP1; 2002-130909/17.

comprising identification of chemically modified gene, useful for diagnosis and treatment of diseases associated with abnormal cytosine methylation -

NAME: , SEQ ID NO 1026; 32pp + Sequence Listing; German.

genes which are modified by a number of human immune system associated can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/allergic bowel diseases. The present sequence is a gene of the invention.

1-10000 M; 2134 A; 453 C; 3328 G; 8413 T; 3 other;

Query Match	4.98;	Score 53.4;	DB 24;	Length 17131,
Best Local Similarity	49.58;	Pred. No. 0.22;		
Matches 138; Conservative	0;	Mismatches 141;		

28 ATACATGTCATTAGAAAGTTCCAAAAAATTTTGTAAACGAAGCTTCCAATTTTTT 87
| | | | | | | | | | | | | | | | | |
6867 ATATTAAGTTCATTAATAATTATTAACGAAAAAAAATTAATAAANCACTTAATTTAT 6806
| | | | | | | | | | | | | | | | | |

[illegible]

RESULT 10
ABK28257

AB2023 / standard; DNA; 8170 BP.

AC ABK28257;
xy

23-APR-2002 (first entry)

transcription associated genomic DNA #66. XX

KW PNA; cysteine methylation state; peptide nucleic acid; PNA-oligomer;
 KW single nucleotide polymorphism; SNP; retroviral infection; gene; ds;
 KW viral infection; Sezary syndrome; haematologic disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; neurological disorder; Niemann-pick disease;
 KW myeloid/plastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angioedema; congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamin disorder; solid tumour.

Unidentified

WO200192565-A2

06-DEC-2001

06-APR-2001; 2001WO-EP03973

06-APR-2000; 2000DE-1019058
07-APR-2000; 2000DE-1019173

01-SEP-2000; 2000DE-1043826

(EPIC-) EPIGENOMICS AG.

Olek A, Plepenbrock C, Berlin K;

WPL; 2002-090046/12.

diseases associated with DNA transcription, e.g. immunological disorders, Werner syndrome, psoriasis, myocardial infarction, solid tumours or cancer -

claim 1; SEQ ID NO 131; 32pp; English.

Chemically pretreated DNA of nucleic acids, which comprises a segment of one of 346 sequences, and of states associated with DNA transcription from one or peptide nucleic acid (PNA)-oligonucleotide. In particular an oligonucleotide to the chemically pretreated DNA of probes associated with DNA transcription. The set of oligomer probes are useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) in a chemically pretreated genomic DNA. The nucleic acids are useful for

CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Wardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polylutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.

SQ Sequence 8170 BP; 2772 A; 69 C; 1274 G; 4055 T; 0 other;

Query Match	4.88	Score 53.2	DB 24	Length 8170
Best Local Similarity	47.88	Pred. No. 0.21		
Matches 154	Conservative 0	Mismatches 168	Indels 0	Gaps 0

QY	665	ATGACCTTTTCTTTTTCACCGCGGTGGATGGAAGCTTTAGTACTTTAGCCAGAGACAAAT	724
Db	311	ATTATTATTATTTTTATATATATGATAAAAAATTTTAAATATATAAAGGGTAAATATAT	370
QY	725	TGATTATATGATATATCCATTATCCAGATATTTATGATATAAATTTACCTGTTAAACATTT	784
Db	371	TGTTTATTAATAGTATAGTATTATTAAGTATTAAGTTGGGTTATATAGTTAAGTATGTTTAAT	430
QY	785	TCAGCATCGACGCTTTCGCAACTTTTGTGTTTAAATTAAGTATTAATTAATTAAGTA	844
Db	431	TATGAAATTTTATTTTATGTTGGAATATATATATGTTTGAATTTTATAAAAAAAAGTA	490
QY	845	TTTAAAGGAGCATTAACGAGCAAAAAAGTATTAACACGAGAAAAAAAAGCATGAG	904
Db	491	GAAAAATTTGAAAAATATTAATATGAAAAATAGCAAAAAAGCGTTTAAATTTATATATATA	550
QY	905	CTCATTTGGTTAGTTAAAGCTTAATAGACAGATTTTATTAATTTTAAATGACGATGATAAC	964
Db	551	ATTAGACATATATAGATTTGAAATATTAATTTTAAATAGAAATTTTTTTGGGGTTATAT	610
QY	965	AATTATATTTTCGACTCTTT	986
Db	611	ATTTTATTTTATTTTATTTT	632

RESULT 11
ABK22782
ID ABK22782 standard; cDNA; 54108 BP.

AC	ABK22782;
XX	
DT	09-APR-2002 (first entry)
YY	

Human high bone mass (HBM) polynucleotide clone #5.

KM Human; mouse; Zmax; HBW: high bone mass gene; lipid regulation; stroke;
KM lipid-associated condition; arteriosclerosis; cardiovascular disease; ss;
KM osteoporosis; atherosclerosis; diabetic atherosclerosis; plaque build-up;
KM neurovascular condition; wound healing; gene therapy; PCR primer; probe;
KM bone development disorder; antiatherosclerotic; cardiovascular;
KM osteopathic; cerebroprotective.

Homo sapiens.

PN WO200192891-A2.

PD 06-DEC-2001.

PF 25-MAY-2001; 2001WO-US16946.

PR 26-MAY-2000; 2000US-0578900.

PA (GENO-) GENOME THERAPEUTICS CORP.
PA (UYCR-) UNIV CREIGHTON SCHOOL MEDICINE.
XX
XX
PI Carulll JP, Little RD, Recker RR, Johnson ML,
XX WPI; 2002-097784/13.
DR
XX

PT Identifying molecules involved in lipid regulation, useful for
PT diagnosing, treating or preventing e.g., arteriosclerosis, comprises
PT identifying a molecule that binds to high bone mass gene or its
PT corresponding wild type gene -

PS Example 2; Page 284-323; 409pp; English
xy

The invention relates to a method for identifying a molecule involved in lipid regulation comprising identifying a molecule that binds to or inhibits binding of a molecule to high bone mass (HBM) or its wild type gene, *zmx1*. Compounds identified by the method are useful for treating, diagnosing, preventing or screening for normal and abnormal lipid-associated conditions, including arteriosclerosis, cardiovascular disease, stroke, and osteoporosis. The compounds may also be used in the treatment or prevention of diabetic atherosclerosis, neurovascular conditions caused by plaque build-up, poor circulation, poor plaque build-up and associated poor wound healing. The methods may be used in gene therapy, pharmaceutical development, and diagnostic assays for bone development disorders. Molecules identified by comparison of *zmx1* and HBM systems can be used as surrogate markers in pharmaceutical development, in diagnosis of human or animal bone disease, and in the treatment of bone diseases. Sequences ABK22776-ABK2411 represent cDNA molecules encoding human *zmx1* and HBM, and PCR primers, probes, linkers and adapters of the invention.

SQ Sequence 54108 BP; 11394 A; 14471 C; 15121 G; 13120 T; 2 other;

Query Match	4.88	Score	53	DB	24	Length	54108
Best Local Similarity	50.28	Pred. No.	0.31				
Matches	157	Conservative	0	Mismatches	155	Indels	1
				Gaps			1

OY	6	ACAATTTTAAATTATTTGGTCATACATGATCGAAGGTCCAAAAATTTTTGGTA	65
Db	27802	ATAAATATAAAAATATATATTTTAAATATATAATATATATAATATAATATATTTAA	27866
OY	66	ACAGAACTCCAAATTTTTTTTTTTATGA - ACAGAAATPACAGATAGAAACTATT	124
Db	27862	AATAAAATTTTAAATTTATATATATATAGAAATATATATATATATATATATATA	27922
OY	125	TGTGTGTGAATGAGTAGTAAATACATTATAGCAAAATTTAAAAAATATATATAGCT	184
Db	27922	TTATTTATATATATATATAAAATATATATATATATATATATATATATATATAT	27982
OY	185	ATACGCCTCAAAGTATGTTATCTAGTGAGGTAAATTAATTAATGCATGGCGCATCAGA	244
Db	27982	ACATACCTTATAGTATATATATTTAAAAATATATCTAATATGATATATTTTATATGATGATNA	28042
OY	245	ATTGGGACACATCGAATAACGCAATTAATAATTTAACTTAATAATTAATAAATTTGAG	304
Db	28042	TAAATACATTTATATAAACACATTTATATATATTTATATATAAAATATATATAAATCTCC	28102
OY	305	TAAATGTGTTTC 317	
Db	28102	AAGTTGCTTTTTC 28114	

RESULT 12
ABK22784

AC ABK22784

DT	09-APR-2002 (first entry)
----	---------------------------

Human high bone mass (HBM) polynucleotide clone #7.

KW Human; mouse; Zmax1; HBM; high bone mass gene; lipid regulation; stroke;
KW lipid-associated condition; arteriosclerosis; cardiovascular disease; ss;
KW osteoporosis; atherosclerosis; diabetic atherosclerosis; plaque build-up;
KW neurovascular condition; wound healing; gene therapy; PCR primer; probe;
KW bone development disorder; antiarteriosclerotic; cardiovascular;
KW osteopathic; cerebroprotective.
XX
OS Homo sapiens.
XX
PN WO200192891-A2.
XX
PD 06-DEC-2001.
XX
PF 25-MAY-2001; 2001WO-US16946.
XX
PR 26-MAY-2000; 2000US-0578900.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX (UYCR-) UNIV CREIGHTON SCHOOL MEDICINE.
XX
PI Carulli JP, Little RD, Recker RR, Johnson ML;
XX
DR WPI; 2002-097784/13.
XX
PT Identifying molecules involved in lipid regulation, useful for
PT diagnosing, treating or preventing e.g., arteriosclerosis, comprises
PT identifying a molecule that binds to high bone mass gene or its
XX corresponding wild type gene -
XX
PS Example 2; Page 350-392; 409pp; English.
XX
CC The invention relates to a method for identifying a molecule involved in
CC lipid regulation comprising identifying a molecule that binds to or
CC inhibits binding of a molecule to high bone mass (HBM) or its wild type
CC gene, Zmax1. Compounds identified by the method are useful for treating,
CC diagnosing, preventing or screening for normal and abnormal
CC lipid-associated conditions, including arteriosclerosis, cardiovascular
CC disease, stroke, and osteoporosis. The compounds may also be used in the
CC treatment or prevention of diabetic atherosclerosis, neurovascular
CC conditions caused by plaque build-up, poor circulation due to plaque
CC build-up and associated poor wound healing. The methods may be used in
CC gene therapy, pharmaceutical development, and diagnostic assays for bone
CC development disorders. Molecules identified by comparison of Zmax1 and
CC HBM systems can be used as surrogate markers in pharmaceutical
CC development, in diagnosis of human or animal bone disease, and in the
CC treatment of bone diseases. Sequences ABK22776-ABK23411 represent cDNA
CC molecules encoding human Zmax1 and HBM, and PCR primers, probes, linkers
CC and adapters of the invention.
XX
SQ Sequence 57273 BP; 12179 A; 15326 C; 15607 G; 14161 T; 0 other;
Query Match 4.8%; Score 53; DB 24; Length 57273;
Best Local Similarity 50.2%; Pred. No. 0.31;
Matches 157; Conservative 0; Mismatches 155; Indels 1; Gaps 1;

QY 6 ACAAAATTAAATTTAGTCAATACATGATGAGAAAGTCCAAAAAATTTGTTA 65
DB 15934 ATAAATTAATAAATATATATATATATATATATATATATATATATATATA 15993
QY 66 ACAGAACTCCAAATTTTTTTTTTATGGA-ACAAGAAATTAACAGATGAACACTAT 124
DB 15994 AATTAATTTATTAATTTATTAAGTAATTAATTAATTAATTAATTAATTAATA 16053
QY 125 TTGTGTGGAATGGAAGTACTATATACATTAAACAATTTAAAAAATTTAAGCCT 184
DB 16054 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 16113
QY 185 ATACGGCCTCAAAAGTATGTTATCTAGAGTGTAATTAATTAATGATGGCATTCAGA 244
DB 16114 ACATACTTAACTATATATATTAATTAATTAATGATGTAATTTTAAAGTATGATATA 16173
QY 245 ATTGGGCAACAAATGAAGGCAATTAATTAATTAATTAATTAATTAATTAATTTGAG 304

DB 16174 TAATATACATTTTAATACACATTTATATATATATATATATATATATATATATAT 16233
QY 305 TAAATGTGTTTC 317
DB 16234 AAGTCTCTTTTC 16246
RESULT 13
ABA82625
ID ABA82625 standard; DNA; 66933 BP.
XX
AC ABA82625;
XX
DT 25-JAN-2002 (first entry)
XX
DE Human HBM gene region b200e21-h_contig4.
XX
KW Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
KW sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
KW antisense therapy; vaccine; bone disorder; Paget's disease;
KW sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX
OS Homo sapiens.
XX
PN WO200177327-A1.
XX
PD 18-OCT-2001.
XX
PF 21-JUN-2000; 2000WO-US16951.
XX
PR 05-APR-2000; 2000US-0543771.
XX 05-APR-2000; 2000US-0544398.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Carulli JP, Little RD, Recker RR, Johnson ML;
XX
DR WPI; 2001-657171/75.
XX
PT New high bone mass (HBM) and Zmax1 genes and proteins useful for
PT modulating bone mass for the treatment of e.g. osteoporosis -
XX
PS Claim 51; Page 308-350; 443pp; English.
XX
CC The present invention describes the human Zmax1 gene and the high bone
CC mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and
CC HBM genes have osteopathic activities. The genes can be used in gene
CC therapy, antisense therapy and in the production of vaccines. They
CC can be used in the diagnosis and treatment of bone disorders including
CC osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
CC dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 66933 BP; 14237 A; 17817 C; 18323 G; 16556 T; 0 other;
Query Match 4.8%; Score 53; DB 22; Length 66933;
Best Local Similarity 50.2%; Pred. No. 0.32;
Matches 157; Conservative 0; Mismatches 155; Indels 1; Gaps 1;

QY 6 ACAAAATTAAATTTAGTCAATACATGATGAGAAAGTCCAAAAAATTTGTTA 65
DB 26974 ATAAATTAATAAATATATATATATATATATATATATATATATATATATA 27033
QY 66 ACAGAACTCCAAATTTTTTTTTTATGGA-ACAAGAAATTAACAGATGAACACTAT 124
DB 27034 AATTAATTTATTAATTTATTAAGTAAATTAATTAATTAATTAATTAATTAATA 27093
QY 125 TTGTGTGGAATGGAAGTACTATATACATTAAACAATTTAAAAAATTTAAGCCT 184
DB 27094 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 27153
QY 185 ATACGGCCTCAAAAGTATGTTATCTAGAGTGTAATTAATTAATGATGGCATTCAGA 244

DB 27154 ACATCTTATAGTATATATTAATAATATGTAATATTTTATTTATATGATATATA 27213
QY 245 ATTGGACACATGAAACGGAATTAAATATTAACCTTTAAATTAATAAATTGGAG 304
DB 27214 TAATATACATTTTAAATACATTTTATTTTATTTATTAATAAATATATATATATCTCC 27273
QY 305 TAAATGTGTTTC 317
DB 27274 AAGTTGCTTTTC 27286

RESULT 14
ABA82623
ID ABA82623 standard; DNA; 72049 BP.
XX ABA82623;
AC
XX
XX
DT 25-JAN-2002 (first entry)
XX
DE Human HBM gene region b527812-h_contig309G.
XX
KW Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
KW sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
KW antisense therapy; vaccine; bone disorder; Paget's disease;
KW sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX
OS Homo sapiens.
XX
PN WO200177327-A1.
XX
PD 18-OCT-2001.
XX
PF 21-JUN-2000; 2000WO-US16951.
XX
PR 05-APR-2000; 2000US-0543771.
PR 05-APR-2000; 2000US-0544398.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX
PI Carulli JP, Little RD, Recker RR, Johnson ML;
XX
DR WPI; 2001-657171/75.
XX
XX
PT New high bone mass (HBM) and Zmax1 genes and proteins useful for
PT modulating bone mass for the treatment of e.g. osteoporosis -
XX
XX
PS Claim 51; Page 258-302; 443pp; English.
XX
CC The present invention describes the human Zmax1 gene and the high bone
CC mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and
CC HBM genes have osteopathic activities. The genes can be used in gene
CC therapy, antisense therapy and in the production of vaccines. They
CC can be used in the diagnosis and treatment of bone disorders including
CC osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
CC dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 72049 BP; 15164 A; 19306 C; 20142 G; 17434 T; 3 other;

Query Match 4.8%; Score 53; DB 22; Length 72049;
Best Local Similarity 50.2%; Pred. No. 0.32;
Matches 157; Conservative 0; Mismatches 155; Indels 1; Gaps 1;

QY 6 ACAATATATATATATAGTCATTAACAATGCAATGAAAGTCCAAAAAATTTTGTGA 65
DB 30562 ATAAATATATAATA 30621
QY 66 ACAGAACTTCCAAATTTTATTTTATGCA-ACAAGAAATACAGATACAAAATCTATT 124
DB 30622 AATAAATA 30681
QY 125 TTGTTGTGCAATGCAATGATATATATATATATATATATATATATATATATATATAT 184

DB 30682 TTAT 30741
QY 185 ATACGGCGCTCAAGATGTTATCTAGTGTGTAATTAATGATGCGGATTCAGA 244
DB 30742 ACATCTTATAGTAT 30801
QY 245 ATTGGACACATGAAACGGAATTAAATATTAACCTTTAAATTAATAAATTGGAG 304
DB 30802 TAATATACATTTTAAATACATTTTATTTTATTTATTAATAAATATATATATCTCC 30861
QY 305 TAAATGTGTTTC 317
DB 30862 AAGTTGCTTTTC 30874

RESULT 15
ABL70459/C
ID ABL70459 standard; DNA; 16228 BP.
XX ABL70459;
AC
XX
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence#175.
XX
KW Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.
XX
OS unidentified.
XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP07471.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154758/20.
XX
XX
PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling -
XX
XX
PS Claim 1; SEQ ID NO 349; 24pp+sequence listing; English.
XX
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligonucleotides for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office.
XX
SQ Sequence 16228 BP; 4652 A; 198 C; 3601 G; 7777 T; 0 other;

Query Match 4.8%; Score 52.6; DB 24; Length 16228;
Best Local Similarity 48.2%; Pred. No. 0.31;
Matches 148; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 47 TCCAAAAAATTTGTTACAGAACTCCAAATTTTTTTTTTTATGGAACAGAAAT 106
 Db 7878 TCTACTATTAACTTATTTTAAAAAAACAACCTTTTCTTTTAAACGAAATA 7819
 QY 107 AACAGTAGAAACTATTTTGTGGAATGGAAGTAGTAATATACATTAAAGCAATTTT 166
 Db 7818 TACCGTAACTAATTTATTTCCCAATTCCTTAATACCTCAATTAACCTAATTAATAACA 7759
 QY 167 AAAAAATTATATAGCCATAAGGCGCTCAAGTATGTTATCTAGTAGTGTAATATA 226
 Db 7758 AAAAAAATAAATAAATAAAGAAATACCTTAATTAATTAATTTACTTCCCTAT 7699
 QY 227 TGCATGTCGATCAGATTCGACACAAATGAAACGAATTAAATATTAACCTTAA 286
 Db 7698 TAAACATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 7639
 QY 287 AATTAATTAATAATTTGAGTAATGTGTTCTGACTATGAGGGGCAAAAAAGACAAAT 346
 Db 7638 AACTATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7579
 QY 347 GCCAAAA 353
 Db 7578 TAAATAA 7572

Search completed: March 29, 2003, 18:32:56
 Job time : 644.917 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:09:40 ; Search time 38.5031 Seconds

(without alignments)
8769.450 Million cell updates/sec

Title: US-09-502-426A-1_COPY_2102_3202

Sequence: 1 aatctacaattatataat.....gagagagagaactagctcc 1101

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	49.2	4.5	3350	US-08-617-860B-3
2	49.2	4.5	6265	US-09-129-112-3
3	46.6	4.2	19124	US-08-487-826B-13
4	46.4	4.2	636	US-08-998-416-1137
5	46	4.2	731	US-08-451-405A-2
6	45.8	4.2	665	US-08-883-795A-36
7	45.8	4.2	51952	US-08-947-823-1
8	45	4.1	20674	US-07-867-106-4
9	44.8	4.1	3138	US-09-641-638-651
10	44.6	4.1	19124	US-08-487-826B-13
11	44	4.0	7218	US-08-232-463-14
12	43.8	4.0	6124	US-08-213-419B-3
13	43.6	4.0	5852	US-07-867-106-2
14	43.2	3.9	660	US-07-991-867B-32
15	43.2	3.9	660	US-08-107-755A-32
16	43.2	3.9	660	US-08-544-332-32
17	43.2	3.9	660	US-09-370-861A-32
18	43.2	3.9	1511	US-07-991-867B-8
19	43.2	3.9	1511	US-08-107-755A-8
20	43.2	3.9	1511	US-08-544-332-8
21	43.2	3.9	1511	US-09-370-861A-8
22	43.2	3.9	4810	US-08-852-629-11
23	43.2	3.9	4838	US-08-852-629-15
24	42.8	3.9	1850	US-08-617-860B-32
25	42.8	3.9	4098	US-08-605-106-4
26	42.4	3.9	470	US-09-020-956-102
27	42.4	3.9	470	US-09-030-607-102

28	42.4	3.9	470	US-09-605-785-102	Sequence 102, App
29	42.4	3.9	470	US-09-439-313-102	Sequence 102, App
30	42.4	3.9	470	US-09-352-616A-102	Sequence 102, App
31	42.4	3.9	470	US-09-232-149A-102	Sequence 102, App
32	42.4	3.9	615	US-08-998-416-186	Sequence 186, App
33	42.4	3.9	5181	US-08-257-073-10	Sequence 10, App1
34	42.2	3.8	1406	US-08-913-842-6	Sequence 10, App1
35	41.8	3.8	658	US-08-998-416-595	Sequence 595, App
36	41.4	3.8	2861	US-08-299-953-1	Sequence 1, App1
37	41.4	3.8	2861	US-08-459-415-1	Sequence 1, App1
38	41.4	3.8	2861	US-09-066-687-1	Sequence 1, App1
39	41.4	3.8	2861	PCT-US95-11231-1	Sequence 1, App1
40	41.4	3.8	3881	US-08-239-953-2	Sequence 2, App1
41	41.4	3.8	3881	US-08-459-415-2	Sequence 2, App1
42	41.4	3.8	3881	US-09-066-687-2	Sequence 2, App1
43	41.4	3.8	3881	PCT-US95-11231-2	Sequence 2, App1
44	41	3.7	1441	US-08-821-994-63	Sequence 63, App1
45	41	3.7	2110	US-09-419-459-1	Sequence 1, App1

ALIGNMENTS

RESULT 1
US-08-617-860B-3/C

Sequence 3, Application US/08617860B

Patent No. 6133506

GENERAL INFORMATION:

APPLICANT: Tytfer, R., Bautor, J., Bothmann, H., Filssak, E.,
Applicant: Tytfer, R., Bautor, J., Bothmann, H., Filssak, E.,
Applicant: M. Iler, A., Schulte, W., Voeltz, M., Walek, J.,
Applicant: Schell, J.,

TITLE OF INVENTION: Promoters

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESSES:

ADDRESS: Steinberg, Raskin & Davidson, P.C.

STREET: 1140 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

APPLICATION NUMBER: US/08/617,860B

FILING DATE: 01-MAR-1996

APPLICATION NUMBER: PCT/EP94/02950

FILING DATE: 05-SEP-1994

APPLICATION NUMBER: DE P4329551.2

FILING DATE: 04-SEP-1993

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3350 base pairs

TYPE: Nucleic acid

STRANDEDNESS: Double stranded

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Brassica napus

IMMEDIATE SOURCE:

LIBRARY: genomic Lambda FIX II

CLONE: Baccaseg10

FEATURE:

NAME/KEY: Startcodon

LOCATION: 2611..2613

FEATURE:

NAME/KEY: CDS

```

; LOCATION: join(2611..2908, 3001..3341)
US-08-617-8608-3

Query Match      4.5%; Score 49.2; DB 3; Length 3350;
Best Local Similarity 48.7%; Pred. No. 0.02;
Matches 190; Conservative 0; Mismatches 198; Indels 2; Gaps 2

OY   6 ACAAATTTATTATATTTTTGTGCAATAACATGCATGAAGTCCAAAATAAATTTTGTTA 65
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   2312 ATAAATGATTAATATTTTATATATATATATATAT -CAGCTATGTAATATATAATATAGAA 2254
OY   66 ACAGAACAATCCAAATTTTTTTTTTTTTTATATGACAACAAGAAATATACAGATAGAAAACTATTT 125
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   2253 ACTAAATTAATTAATTAATTAATGTTTTTAAACAATAATATAGTTGGGGA-TGTAT 2195
OY   126 TGTGTGGAAATGGAAGTAGTAATATACATTAAGCAAAATTTTAAAAAATTAATATACCGCTA 185
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   2194 ATATTTTATGTTTTAAAAACATTTAGAAAATATTTAATTTTAACTTTTAAATTTTATTTTC 2135
OY   186 TAGCGGCCCAAGTNGTACTAGTAGTGTAATTAATAATGCAATGGCCATTCAGAA 245
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   2134 ATTGACATATCAAAATATCAAAAAAATGTTAGACTATTTAATAATTTTTCACAAAGT 2075
OY   246 TTGGAGCAACAATGAAACGCAATTAAATATTAATTAATTAATTAATTAATTAATTTGACT 305
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   2074 AAGATTTATGATATATATATATATATATATATATATATATATATATATATATATAT 2015
OY   306 AAATGTGTTTTCTGACTATGAGGGGCAAAAAAAGAACAAATCCAAAGCTACGGGTTT 365
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   2014 GTGAAATTAATTAATTTAAACTTTAAATATACAAAATTAACCAAAAATTAAGAATCATTTGATA 1955
OY   366 GACGTCCAGTCGGTAAATCTAATATAC 395
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB   1954 AATGTCAATACAGAAATTAACCTAACACC 1925
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
US-09-129-112-3/c
Sequence 3, Application US/09129112
Patent No. 6465716
GENERAL INFORMATION:
APPLICANT: Etzler, Marilyn E.
APPLICANT: Murphy, Judith B.
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: A No. 6465716 Factor Binding Protein From Legume Roots
FILE REFERENCE: 023070-0798100S
CURRENT APPLICATION NUMBER: US/09/129.112
PRIOR FILING DATE: 1998-08-04
PRIORITY APPLICATION NUMBER: US 08/907,226
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 6265
TYPE: DNA
ORGANISM: Dolichos biflorus
FEATURE:
OTHER INFORMATION: genomic sequence of NBP46 (DB46)
NAME/KEY: exon
LOCATION: (633)..(944)
NAME/KEY: intron
LOCATION: (945)..(1022)
NAME/KEY: exon
LOCATION: (1023)..(1151)
NAME/KEY: intron
LOCATION: (1152)..(1559)
NAME/KEY: exon
LOCATION: (1560)..(1616)
NAME/KEY: intron
LOCATION: (1617)..(1697)
NAME/KEY: exon
LOCATION: (1698)..(1790)
US-09-129-112-3

```

```

Query Match          4.5%: Score 49.2; DB 4; Length 6265;
Best Local Similarity 53.1%: Pred. No. 0.023;
Matches 173; Conservative 0; Mismatches 148; Indels 5; Gaps 3;

QY 15 TAATATTAGTCAATTAACAAATGATAGAAAGTTCACAAAAAATTTTGTATACAGAAACT 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2936 TATTTATAGAGTGTTTACTATTCAATTAACAAATTAATGCAGTATGAAGTTTACTG-AACT 2878

QY 75 TCAGAAATTTTTTTTTTTTTATGACACAGAAATTAACAGATGAAAACATTTTGTGTGGA 134
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2877 TCTTTTTTTTTTTTTTTTTTATTTGAAAAAGTTTGAATAATGTCAAAATTTAAATATAGTGA 2818

QY 135 ATGCAAGTAGTAATATACATTA--GCAATTTTAAAAAATTTATATAGCTTATAGCGC 192
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2817 ATATTTATAGAAAAATGCAATTAATGAAAACTGAAAAAATTAATTAATTAATAGTAGA 2758

QY 193 TCAAGTAGTGTTCATAGTAGTGTAATTAATA--ATGATGTGTGCATTCAGAAATTTGGC 250
   ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2757 TCACACACAGTGTGAAATTTAATTTAAACATATATTAATAATTAACAAATACATTTTAT 2698

QY 251 ACAACAAATGAAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGATAATG 310
   || ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2697 CTATTAATAATAAATAATTAATTTAATAATAATAATAATTAATAATTAATTAATGATATGCTA 2638

QY 311 TGTTCGACTATTGAGGGGCAAA 336
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2637 TTTTATATATCATTAAGCAGACAA 2612

RESULT 3
US-08-487-826B-13
: Sequence 13, Application US/08487826B
: Patent No. 5993827
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chlunis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,826B
: FILING DATE: 10-SEP-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelson, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121.001CP1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 19124 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear

```

```

; MOLECULE TYPE: CDNA
;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-487-826B-13

```

Query Match	4.2%	Score 46.6;	DB 2;	Length 19124;
Best Local Similarity	47.4%	Pred. 0.12;	Indels 0;	Gaps 0;
Matches 139;	Conservative	0;	Mismatches 154;	

QY 9 AATATTATTTTGTGTCATACAGCATAGAAGAAAGTGCCAAAAAATTTGGTRACA 68
| | | | | | | | | | | | | | | | | | | | | |
Db 15735 AAAATAATATATCATTAAAATTAATAAAAAAAAAATGTTTTAAAAAATAATADCA 15794
| | | | | | | | | | | | | | | | | | | | | |
QY 69 GAACCTCCAATTTTTTTTTTATVGAACAGAAATACAGATAGAAACTATTGTT 128
| | | | | | | | | | | | | | | | | | | | | |
Db 15795 TAAAAATAAAAAAATTTTATTTTAAATAAAAAATAATATATTAATTAATTAAN 15855
| | | | | | | | | | | | | | | | | | | | | |
QY 129 TGCGAATGGAAGTAGTATATATACCTTAGCAAAATTTTAAAAATTTATACCGTATAC 188
| | | | | | | | | | | | | | | | | | | | | |
Db 15855 AAAAAAAT 15914
| | | | | | | | | | | | | | | | | | | | | |
QY 189 GCGCTCAAAGTATGTTATCTAGTAGGTGTAATTAATGCATGTCAGTACAGATTG 248
| | | | | | | | | | | | | | | | | | | | | |
Db 15915 AAATAAAAAATATAAAAAATTTTAAATTAATTAATAAAAAATTAATTAATTAATKAT 15974
| | | | | | | | | | | | | | | | | | | | | |
QY 249 GGACACATGAAAAAGCATTAATAATATACCTTAATAATTAATAATTTT 301
| | | | | | | | | | | | | | | | | | | | | |
Db 15975 GCACATATCATACAT 16027

RESULT 4
US-08-998-416-1137/c

Sequence 1137, Application US/08998416
 Patent No. 6239264
 GENERAL INFORMATION:
 APPLICANT: Philippsen, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendlandt, Jurgens
 APPLICANT: Knechtle, Philipp
 APPLICANT: Reibschung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
 TITLE OF INVENTION: AND USES THEREOF
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESS:

ADDRESSSEE: NO. 6239264artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NO. 62392644th Carolina
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998, 416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meliss J. Timochy
 REGISTRATION NUMBER: 38, 241
 REFERENCE/DOCKET NUMBER: PE/5-30306/R/CGCI976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 1137:
 SEQUENCE CHARACTERISTICS:

```

?      LENGTH: 636 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      ORIGINAL SOURCE:
?      ORGANISM: PAGI692RP
?
US-08-998-416-1137

```

Query Match	4.28;	Score 46.4;	DB 4;	Length 636;
Best Local Similarity	48.88;	Pred. NO. 0.055;		
Matches 125; Conservative	0;	Mismatches 131;	Indels 0;	Gaps 0

QY	46	TTCGAAAAAAATTTTGTGAACAGAACTCCAAATT	TTTTTTTATTGACAAGAAGA	10
Dd	636	TTTTAATACTATTTTTTAACACACTCTTTTATTA	AATTTTAAATTAATAATATGAT	57
QY	106	TAAACAGATGAAGAACTATTTTGTGGAGTAG	AGTAGTAATACATTAAGCAAAATT	16
Dd	576	AAAATATATATTAATTAATTAATTAATTAAT	TAAATTAATTAATTAATTAATTAATTAATTAAT	51
QY	166	TAAAAATATATTAAGCCTATACCOCGC	CMAAGATGTATCTAGTAGGTGAATTAATA	22
Dd	516	TATTAATATATTTAAAAATTTAATTAAGAAAT	TTAAAGTTAAATTTTAAATTAATA	45
QY	226	ATGCATGTGCGATTCAGAAATGGACACATGA	AGAAACGAATTTAAATATTAACCTTA	28
Dd	456	TTCCTATAAAAAGATTAATAATATTAATCAT	CAACTAATATTTTAAAAAATCAATTTTATA	39
QY	286	AAATTAATAAAAATTT	301	
Dd	396	ATRAAAATTAATTAATT	381	

RESULT 5
US-08-451-405A-2

Sequence 2, Application US/08451405A
Patent No. 5736358
GENERAL INFORMATION:
APPLICANT: FASEL, NICOLAS JOSEPH
APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE
TITLE OF INVENTION: DIGTYOSLELD EXPRESSION VECTOR AND
TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:

ADDRESS: THE WEBB LAM FILM
 STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE
 CITY: PITTSBURGH
 STATE: PENNSYLVANIA
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 15219-1818
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" FLOPPY DISK
 COMPUTER: Midwest Micro 486-50
 OPERATING SYSTEM: DOS
 SOFTWARE: WORDPERFECT 6.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/451,405A
 FILING DATE: 26-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/965,273
 FILING DATE: 15-JAN-1993
 INFORMATION FOR SEO ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 731
 TYPE: NUCLEIC ACID
 STRANDEDNESS: SINGLE
 TOPOLOGY: UNKNOWN
 US-08-451-405A-2

Query Match	4.28;	Score 46;	DB 1;	Length 731,
Best Local Similarity	46.38;	Pred. No. 0.07;		

Matches 151: Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```
OY 15 TAAATTTAGTCAATACAGATGAGAAAGTCCAAAAAATTTGTTACAGAAACT 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22 TTAATATTTGGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 81
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 75 TCCAAATTTTATTTTATGAGACAGAAATACAGATGAGAACTTTTGTGTGGA 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 82 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 135 ATGAGAGATGATATATACATTAAGCAATTTTAAAAATTTATATTAAGCTATACGCTC 194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 GGAGTTTTTATTTGATTTTAAATATATATATATATATATATATATATATATAT 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 195 AAGATATGTTATCTAGTGTAGTGTATTAATATATATATATATATATATATATAT 254
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 ACGATATATGATATGATATATATATATATATATATATATATATATATATATAT 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 255 CAATGAAACGGAATTAATATATATATATATATATATATATATATATATATATAT 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 262 CAACCAAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTT 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 315 TTCTGACTATGAGGGGCAAAAAA 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 322 ATCAAAAAAAGGTATTTAAAGAA 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 6

US-08-883-795A-36

Sequence 36, Application US/08883795A

Patent No. 5985607

GENERAL INFORMATION:

APPLICANT: Delcive, Genevieve

APPLICANT: Awang, Gregor

TITLE OF INVENTION: Recombinant DNA Molecules and Expression

TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSEE: BERESKIN & PARR

STREET: 40 King Street West

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5H 3Y2

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/883,795A

FILING DATE: 27-JUN-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Gravelle, Micheline

REGISTRATION NUMBER: 40,261

REFERENCE/DOCKET NUMBER: 7841-062

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 364-7311

TELEFAX: (416) 361-1398

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 665 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE: Homo sapiens

ORGANISM: Homo sapiens

IMMEDIATE SOURCE: ClONE: Rh 32

US-08-883-795A-36

Query Match 4.2%; Score 45.8; DB 2; Length 665;

Best Local Similarity 47.7%; Pred. No. 0.076;

Matches 200; Conservative 0; Mismatches 212; Indels 7; Gaps 2;

```
OY 586 TTTATTAATTTTAAACACACAGCTGATGATGATGATGATGATGATGATGATGAT 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 TTTTATTAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 646 GATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 97 ATTATTAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 706 ACTATTAGCAGACAGACATGATGATGATGATGATGATGATGATGATGATGATGAT 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 157 A-TTTATTAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 766 AATATGCTGTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 AATTTTATTAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 826 AGTTTATTAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 AGTTTATTAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 886 AGAAACAAAGCCATGAGACCTGATGATGATGATGATGATGATGATGATGATGAT 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 AGTTTATTAATTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 946 TTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1004
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 TTAATATGTTTATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7

US-08-947-823-1

Sequence 1, Application US/08947823

Patent No. 6114605

GENERAL INFORMATION:

APPLICANT: Williamson, Valerie M.

APPLICANT: Kaloshian, Isigouhi

APPLICANT: Yaghoobi, Jafar

APPLICANT: Bodeau, John

TITLE OF INVENTION: Procedures and Materials for Confering

TITLE OF INVENTION: Pest Resistance in Plants

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/947,823

FILING DATE: 09-OCT-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/18802

FILING DATE: 09-OCT-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/028,191

FILING DATE: 10-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Bastian, Kevin L.

REGISTRATION NUMBER: 34,774

REFERENCE/DOCKET NUMBER: 023070-070210US

TELECOMMUNICATION INFORMATION:

OTHER INFORMATION: 10-513-250	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 2844	
OTHER INFORMATION: 10-513-262	: polymorphic base C or T
NAME/KEY: allele	
LOCATION: 2934	
OTHER INFORMATION: 10-513-352	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 2947	
OTHER INFORMATION: 10-513-365	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 3802	
OTHER INFORMATION: 12-206-81	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 4062	
OTHER INFORMATION: 10-343-231	: deletion of C
NAME/KEY: allele	
LOCATION: 4088	
OTHER INFORMATION: 12-206-366	: polymorphic base C or T
NAME/KEY: allele	
LOCATION: 4109	
OTHER INFORMATION: 10-343-278	: polymorphic base C or T
NAME/KEY: allele	
LOCATION: 4170	
OTHER INFORMATION: 10-343-339	: polymorphic base G or T
NAME/KEY: allele	
LOCATION: 5903	
OTHER INFORMATION: 10-346-23	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 6019	
OTHER INFORMATION: 10-346-141	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 6141	
OTHER INFORMATION: 10-346-263	: polymorphic base G or C
NAME/KEY: allele	
LOCATION: 6183	
OTHER INFORMATION: 10-346-305	: polymorphic base C or T
NAME/KEY: allele	
LOCATION: 6338	
OTHER INFORMATION: 10-347-74	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 6375	
OTHER INFORMATION: 10-347-111	: polymorphic base G or C
NAME/KEY: allele	
LOCATION: 6429	
OTHER INFORMATION: 10-347-165	: polymorphic base C or T
NAME/KEY: allele	
LOCATION: 6467	
OTHER INFORMATION: 10-347-203	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 6484	
OTHER INFORMATION: 10-347-220	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 6534	
OTHER INFORMATION: 10-347-271	: polymorphic base A or T
NAME/KEY: allele	
LOCATION: 6611	
OTHER INFORMATION: 10-347-348	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 7668	
OTHER INFORMATION: 10-348-391	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 8608	
OTHER INFORMATION: 10-349-47	: polymorphic base C or T
NAME/KEY: allele	
LOCATION: 8658	
OTHER INFORMATION: 10-349-97	: polymorphic base A or G
NAME/KEY: allele	
LOCATION: 8703	
OTHER INFORMATION: 10-349-142	: polymorphic base G or C
NAME/KEY: allele	
LOCATION: 8777	
OTHER INFORMATION: 10-349-216	: deletion of CTG

```

NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

```

Query Match	4.1%	Score 45	DB 4	Length 20674
Best Local Similarity	45.9%	Pred. No. 0.27		
Matches 153	Conservative 0	Mismatches 180	Indels 0	Gaps
QY 1	AATCTACAATATATTAATATTTAGTCAATTAACAAATGATAGAAAGTCCAAAAAATTT	60		
Db 11391	AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	11332		
QY 61	TGTTACGAAACCTCCAAATTTTTTTTTTTATGGAACAGAAATTAACAGTACAGAAC	120		
Db 11331	ATATTAATTTAATTAATTAATTAAGTTAAATTAATTAATTAATTAATTAATTAATTAAT	11272		
QY 121	TATTTTGGTGGAGATGAGAGTAAATATACATTAACCAATTTTAAATAATTAATTA	180		
Db 11271	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT	11212		
QY 181	GCCATATACGCCCTCAAGATGTTATCTAGTAGTGTAATTAATTAATGATGCTGCAT	240		
Db 11211	AAATTTTAAATTTAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT	11522		
QY 241	CAGAAATGCGACACAAATGAAACGGAATTAATAATTTAACTTTAAATTAATTAATTT	300		
Db 11151	TAAATTTAAATTTAAATTTTAAATTTTAAATTTAAATTTAAATTTTAAATTTAAATTTAT	11092		
QY 301	TGAGTAATGTGTTTTCGACTATGAGGCGCA	333		
Db 11091	TAAATATTTTCCCTTAGCTATTTAAAAAAGGCA	11059		

RESULT 9
 US-07-867-106-4/C
 Sequence 4, Application US/07867106
 Patent No. 5389526
 GENERAL INFORMATION:
 APPLICANT: Slade, Martin B
 APPLICANT: Chang, Andy C M
 APPLICANT: Williams, Keith L
 TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
 TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526/US
 STREET: One Liberty Place 46th Floor
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19103
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

Query Match	4.08;	Score 44;	DB 1;	Length 7218;
Best Local Similarity	7.88;	Pred. NO. 0.35;		
Matches 20;	Conservative 138;	Mismatches 98;	Indels 0;	Gaps 0

RESULT 12
US-08-213-419B-3/C

Sequence 3, Application US/08213419B
Patent No. 6333406
GENERAL INFORMATION:
APPLICANT: Inselburg, J. et al.
TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
FILE OF INVENTION: AND USES THEREFOR
FILE REFERENCE: J11-0020NCP
CURRENT APPLICATION NUMBER: US/08/213,419B
CURRENT FILING DATE: 1994-03-14
PRIOR APPLICATION NUMBER: US 07/870,506
PRIOR FILING DATE: 1992-04-17
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 6124
TYPE: DNA
ORGANISM: Plasmodium falciparum
FEATURE:
NAME/KEY: CDS
LOCATION: (2407)..(2439)
NAME/KEY: CDS
LOCATION: (2359)..(3404)

Query Match	4.0%;	Score 43.8;	DB 4;	Length 6124;
Best Local Similarity	48.2%;	Pred. No 0.37;		
Matches 123;	Conservative 0;	Mismatches 132;	Indels 0;	Gaps 0

RESULT 13
US-07-867-106-2

Sequence 2, Application US/0/86/106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz MacKiewicz & No. 5389526rls
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single

[illegible]

RESULT 14
 US-07-991-867B-32
 ; Sequence 32, Application US/07991867B
 ; Patent No. 5476781
 ;
 GENERAL INFORMATION:
 ; APPLICANT: Moyer, Richard W.
 ; APPLICANT: Hall, Richard L.
 ; APPLICANT: Grudl, Michael F.
 ; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
 ; NUMBER OF SEQUENCES: 66
 ;
 CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: David R. Saliwanchik
 ; STREET: 2421 N.W. 41st Street, Suite A-1
 ; CITY: Gainesville
 ; STATE: FL
 ; COUNTRY: USA
 ; ZIP: 32606
 ;
 COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/991,867B
 ; FILING DATE: 12-DEC-1992
 ; CLASSIFICATION: 435
 ;
 PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO 92/14818
 ; FILING DATE: 12-FEB-1992
 ;
 PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/827,685
 ; FILING DATE: 30-JAN-1992
 ;
 PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/657,584
 ; FILING DATE: 19-FEB-1991
 ;
 ATTORNEY/AGENT INFORMATION:
 ; NAME: Saliwanchik, David R.

Query Match	3.9%	Score 43.2;	DB 1;	Length 660;
Best Local Similarity	44.4%;	Pred. No. 0.29;		
Matches 174;	Conservative	0;	Mismatches 319.	r=0.3

QY	692	ATGAAAGTATTAGACTATTAGCCAGCAGACAAATTGATTATTAAGATAATATCCATTAAACCAT	751
Db	88	ATTAAATTTAATATTATTAAAAAAAATTAGTTAAATTTGGAGAATATGCATATTAAATATTAN	147
QY	752	GATATTTCATGTAAATAAATAGCTTTAAACAATTTCAGCATCCGAGCTTCTCGCACTTT	811
Db	148	GATAAATAAATTTTTAAATATAATATTCAGAAAAATTTAAAGTTATATATTTCCAATTTA	207
QY	812	GTTTTTAATTTTAGACTTTTATTAATAAATAAGTTATTAAGAGCATACGAGCAACANA	871
Db	208	AATATTATTATTTTAAATTTTAACTTTTAAACAAAATTTAAACATTTTAGATATTNCT	267
QY	872	AGTATGAACACGAGAACAAAGCCATGAAAGTCATGTTAGTTTAAAGTTAATAG	931
Db	268	TATTAACAAAAATGACAAATTAGTAAVATNTATACATCCACATTTCTAATGAATTTTAAAT	327
QY	932	AAGATTTTATTAAATTTTATATGACATATATAACAATTATTTCTGCACTTTAAAC	991
Db	328	TGTGATCATGTAAATATATAATGACTATATATTTATTAATAATTAGTAAATTTAAAAA	387
QY	992	CCCCCTTACAAACGAAGCTCCCTTTTTCAGTGAAGTCGATGCCCAACTTAAGAC	1051
Db	388	TTAAATAAATATCTAAAAAAAATTTGGTAACCTTTATAATATGTTTTCCATATAGATATGTT	447
QY	1052	AAAGCATTTAGAAAGAGAGAAAGTGAGTGAGAA	1083
Db	448	GAGTTAAATATGATGCATCATACAAATATAAAGA	479

RESULT 15
 US-08-107-755A-32
 : Sequence 32, Application US/08107755A
 : Patent No. 5721352
 :
 : GENERAL INFORMATION:
 : APPLICANT: Moyer, Richard W.
 : APPLICANT: Hall, Richard L.
 : APPLICANT: Gruidl, Michael E.
 : TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
 : NUMBER OF SEQUENCES: 40
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: David R. Saliwanchik
 : STREET: 2421 N.W. 41st Street, Suite A-1
 : CITY: Gainesville
 : STATE: Florida
 : COUNTRY: U.S.A.
 : ZIP: 32606
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/107,755A
 : FILING DATE: 19-AUG-1993
 : CLASSIFICATION: 435

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:25:26 ; Search time 83.8565 Seconds

(without alignments)
11172.915 Million cell updates/sec

Title: US-09-502-426a-1_COPY_2102_3202

Perfect score: 1101
Sequence: 1 aactacaataatataat.....gagagagagaactagctcc 1101Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.4	4.9	516	10	US-09-960-352-5785 Sequence 5785, Ap
2	53.2	4.8	431	10	US-09-960-352-5558 Sequence 5558, Ap
3	52.2	4.7	377	10	US-09-960-352-7419 Sequence 7419, Ap
4	52.2	4.7	446	10	US-09-960-352-3400 Sequence 3400, Ap
5	49.2	4.5	6265	10	US-09-129-112-3 Sequence 3, Appl1
6	48.6	4.4	53332	9	US-10-224-562-3 Sequence 3, Appl1
7	48.6	4.4	53332	10	US-09-801-861-3 Sequence 3, Appl1
8	48.4	4.4	406	10	US-09-960-352-10265 Sequence 10265, A
9	47.8	4.3	294	10	US-09-960-352-4637 Sequence 4637, Ap
10	47.4	4.3	419	10	US-09-960-352-11234 Sequence 11234, A
11	46.2	4.2	1713	9	US-09-938-842A-4756 Sequence 4756, Ap
12	46.2	4.2	32463	9	US-09-960-352-573 Sequence 573, App
13	45.8	4.2	428	10	US-09-938-842A-3333 Sequence 3333, Ap
14	45.6	4.1	1109	9	US-09-960-352-2819 Sequence 2819, Ap
15	45.6	4.1	413	10	US-09-960-352-1036 Sequence 1036, Ap
16	44.2	4.0	344	10	US-10-072-349-252 Sequence 252, App
17	44.2	4.0	4187	9	US-09-764-855-252 Sequence 252, App
18	44.2	4.0	376	10	US-09-960-352-5087 Sequence 5087, Ap
19	44.0	4.0			

C	20	44	4.0	513509	9	US-09-754-853A-4	Sequence 4, Appl1
C	21	44	4.0	640681	10	US-09-790-988-1	Sequence 1, Appl1
C	22	43.8	4.0	4442	10	US-09-960-352-12911	Sequence 12911, A
C	23	43.8	4.0	2000	9	US-09-938-842A-3251	Sequence 3251, Ap
C	24	43.6	4.0	393	10	US-09-960-352-4582	Sequence 4582, Ap
C	25	43.4	3.9	2000	9	US-09-938-842A-4575	Sequence 4575, Ap
C	26	43.2	3.9	416	10	US-09-960-352-4584	Sequence 4584, Ap
C	27	42.8	3.9	640681	10	US-09-790-988-1	Sequence 1, Appl1
C	28	42.6	3.9	214	10	US-09-960-352-7907	Sequence 7907, Ap
C	29	42.6	3.9	2000	9	US-09-938-842A-4090	Sequence 4090, Ap
C	30	42.4	3.9	351	9	US-09-803-719-1436	Sequence 1436, Ap
C	31	42.4	3.9	424	10	US-09-960-352-11218	Sequence 937, App
C	32	42.4	3.9	437	10	US-09-969-373-937	Sequence 10262, A
C	33	42.4	3.9	451	10	US-09-960-352-10262	Sequence 102, App
C	34	42.4	3.9	470	9	US-09-232-880-102	Sequence 102, App
C	35	42.4	3.9	470	9	US-10-012-896-102	Sequence 102, App
C	36	42.4	3.9	470	9	US-09-895-793-102	Sequence 102, App
C	37	42.4	3.9	470	9	US-09-895-814-102	Sequence 102, App
C	38	42.4	3.9	470	10	US-09-759-143-102	Sequence 102, App
C	39	42.4	3.9	470	10	US-09-780-666-102	Sequence 102, App
C	40	42.4	3.9	470	10	US-09-030-606-102	Sequence 102, App
C	41	42.4	3.9	470	10	US-09-822-827-102	Sequence 102, App
C	42	42.4	3.9	470	10	US-09-115-453-102	Sequence 102, App
C	43	42.4	3.9	1045	9	US-09-938-842A-3089	Sequence 3089, Ap
C	44	42.4	3.9	2000	9	US-09-938-842A-4902	Sequence 4902, Ap
C	45	42.4	3.9	335913	9	US-09-754-853A-2	Sequence 2, Appl1

ALIGNMENTS

```
RESULT 1
US-09-960-352-5785/C
; Sequence 5785, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5785
; LENGTH: 516
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (76) (90)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1
US-09-960-352-5785
```

Query Match 4.9%; Score 54.4; DB 10; Length 516;
Best local similarity 48.7%; Pred. No. 0.09; Mismatches 156; Indels 0; Gaps 0;

```
Matches 148; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
QY 5 TACAAATTATTAATTAGTCATTAACATGCATGAGAAAGTTCCAAAAAATTTGTT 64
DB 460 TAAAAAATAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 401
QY 65 AAGAGAACTCCAAATTTTATTTTATGGAAGAAATATACAGATGAAAACTATT 124
DB 400 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 341
QY 125 TTGTTGTGAGATGAGATAGATATATACATTAAGCAATTTTAAATAATTAAGCC 184
DB 340 TTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 281
```

[illegible]

```

RESULT 2
US-09-960-352-5558/c
: Sequence 5558, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengping
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21.10298/C
: CURRENT APPLICATION NUMBER: US/09/960.352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 5558
: :
: : LENGTH: 431
: : TYPE: DNA
: : ORGANISM: Bos taurus
: : OTHER INFORMATION: Clone ID: 24-LIB3057-024-Q1-K1-F7
: US-09-960-352-5558

```

Query Match	4.88;	Score 53.2;	DB 10;	Length 431;
Best Local Similarity	46.18;	Pred. No. 0.15;		
Matches 178;	Conservative 0;	Mismatches 208;	Indels 0;	Gaps 0;
QY 16	AATATTTGTCAAATACAAATGCATGAGAAAGTCCAAAAAAATTTGTAAACAGAACTT	75		
Db 429	AAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATTTAAT	370		
QY 76	CCAATTTTTTTTTTGTGACAGAAATACAGATAGAAAATCTATTGTGTGGAA	135		
Db 369	ATAAAAATTAATTAATAAAAAATATRAAAAAAATTTAAAAAATTAATTAAT	310		
QY 136	TGGAAGTAGTAATATACATTAAAGCAATTTTAAAAAATATATTAAGCCTATACGGCTCA	195		
Db 309	AATAAAAATATAAAAACAATAAAAAATAAAAAATAAAAAATTAATTAATAAAAAATAAAAA	250		
QY 196	AAGTATGTACTAGTAGTGTATTAATTAATGCATGCTGGCATTCAGAAATTTGGGACAC	255		
Db 249	AAAAATTTAAAAATAAAAAATAAAAAATAAAAAATAAAAAATTAATTAATAA	190		
QY 256	AATGAAACGGAATTTAAATATTAACCTTTAAATTAATTAATTAATTTGATGAATGTGTTT	315		
Db 189	AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	130		
QY 316	TCTGACTATTGAGGGGCAAAAAAAGACAAATGCCAAAAAGCTACGGGTTTGACTGTCAG	375		
Db 129	AAAAAATTTAAAAATAAAAAATAAAAAATAAAAAATTAATTAATTAATTAATTAATTAAT	70		
QY 376	TTGCGTAAATATCTAATTAACCTGTC	401		
Db 69	CAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	44		

RESULT 3
US-09-960-352-7419/c
; Sequence 7419, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:

```

? APPLICANT: Warren, Wesley C.
? APPLICANT: Tao, Nengbing
? APPLICANT: Byatt, John C.
? APPLICANT: Mathialagan, Nagappan
? TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
? TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
? FILE REFERENCE: 16511.006/37-21(10298)C
? CURRENT APPLICATION NUMBER: US/09/960,352
? CURRENT FILING DATE: 2001-09-24
? NUMBER OF SEQ ID NOS: 15112
? SEQ ID NO 7419
? LENGTH: 377
? TYPE: DNA
? ORGANISM: Bos taurus
? OTHER INFORMATION: Clone ID: 32-LIB3057-025-Q1-K1-H11
US-09-960-352-7419

Query Match          4.7%; Score 52.2; DB 10; Length 377;
Best Local Similarity 49.5%; Pred. No. 0.22;
Matches 161; Conservative 0; Mismatches 163; Indels 1; Gaps 1;

```

[illegible]

```

RESULT 4
US-09-960-352-3400/C
: Sequence 3400 Application US//09960352
: Patent No. US20020137159A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengping
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21(10288)C
: CURRENT APPLICATION NUMBER: US/09/960,352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 3400
: LENGTH: 446
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 15-LIB3058-052-01-K1-D11
: US-09-960-352-3400

```

Query Match	4.7%	Score 52.2;	DB 10;	Length 446;
Best Local Similarity	46.7%;	Pred. No. 0.23;		
Matches 165;	Conservative 0;	Mismatches 188;	Indels 0;	Gaps 0;
1 AATCTACAAATTTATTAATTAGTCATTAACAAATGCATAGAAACTTCCAAAAAATTTT 60				

[illegible]

```

RESULT 5
US-09-129-112-3/c
/ Sequence 3, Application US/09129112
/ Patent No. US20020019995A1
/ GENERAL INFORMATION:
/ APPLICANT: Murphy, Marilyn E.
/ APPLICANT: Etzler, Judith B.
/ APPLICANT: The Regents of the University of California
/ TITLE OF INVENTION: A No. US20020019995A1 Factor Binding Protein From Legume Roots
/ FILE REFERENCE: 023070-079810US
/ CURRENT APPLICATION NUMBER: US/09/129,112
/ CURRENT FILING DATE: 1998-08-04
/ PRIOR APPLICATION NUMBER: US 08/907,226
/ PRIOR FILING DATE: 1997-08-06
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 3
/ LENGTH: 6265
/ TYPE: DNA
/ ORGANISM: Dolichos biflorus
/ FEATURE:
/ OTHER INFORMATION: genomic sequence of NBP46 (DB46)
/ NAME/KEY: exon
/ LOCATION: (633)..(944)
/ NAME/KEY: intron
/ LOCATION: (945)..(1022)
/ NAME/KEY: exon
/ LOCATION: (1023)..(1151)
/ NAME/KEY: intron
/ LOCATION: (1152)..(1559)
/ NAME/KEY: exon
/ LOCATION: (1560)..(1616)
/ NAME/KEY: intron
/ LOCATION: (1617)..(1697)
/ NAME/KEY: exon
/ LOCATION: (1698)..(1790)
US-09-129-112-3

Query Match          4.58; Score 49.2; DB 10; Length 6265;
Best Local Similarity 53.18; Pred. No. 2.4;
Matches 173; Conservative 0; Mismatches 148; Indels 5; Gaps 3;

OY      15 TAAATTTAGCAATTAACAATGCATAGAAAAGTCCAAAAAAAATTTGTTAACAAGACT 74
DB      2936 TATTTTATAGAGTTTACTACTATTCACAAATAAACATTAATGCAGATATGAGATTCTG-AACT 2878
OY      75 TCCAAATTTTTTTTTTTTTTATGAGAACAAAGAAATTAACAGATAGAAAACTATTTGGTGGA 134
DB      2877 TCTTTTTTTTTTTTTTTTTATGAGAAATGTTTGAATATTTGCAAAATATTTAATATAGTGA 2818

```

Dy 135 ATGGAAGTAGTAATATACTACATTAA - GCAATTTTAAAAAATAATATATAACCCTATACGGGC 192
Db 2817 AATATTATACAAAATAATGCCAATTAATTGTAAACTAAAAAATAATTAATTATAAGTAAGA 2758

Dy 193 TCAAAGTATGTTATTCCTCGTGCGTAAATATA - ATGCAITGGTGCATTCGAATTGGG 250
Db 2757 TCACAACAAGTTTGAATTTAATTTTTAAACTATATTTAATATTTCACAANAATCATTTAT 2698

Dy 251 ACAACAATGAAAAAGCAATTTAAATTAATCCCTTTAAATTAATAATAAAATTGAGTAATG 310
Db 2697 CTATAAATAATAAAAATTAATTTATATAATAATAATTAATTTATATAAGTATAGTATATGTA 2638

Dy 311 TGTTTCTGCACCTTAGGGGCAAA 336
Db 2637 TTTTATATTTCATTAATCACAGACAA 2612

```

RESULT 6
US-10-224-562-3/c
; Sequence 3, Application US/10224562
; Publication No. US2003002229A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001098D1
; CURRENT APPLICATION NUMBER: US/10/224,562
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 53332
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-224-562-3

Query Match          4.4%; Score 48.6; DB 9; Length 53332;
Best Local Similarity 49.0%; Pred. No. 7;
Matches 129; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY  47  TCCAAAAAAATTTTGTGTACAGAAACTCCAAATTTTTTTTTTTTATAGCAAGAAAT 106
      ||| ||||| ||| | | | | | | | | | | | | | | | | | | | | | | |
Db  31671 TCAAAAAAAAATAATTCATATATATATATATATATATATATATATATATATATAT 31612

QY  107 AACAGATAGAAACATATTTTGTGTGGAATGGAAGTACTAATATATACATTATAGCAAAATTTT 166
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  31611 AATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 31552

QY  167 AAAAAATTTATTAAGCCATATACGCCCTCAAGTATGTTATCTAGTAGTGATTAATTA 226
      ||| ||| ||| | | | | | | | | | | | | | | | | | | | | | | |
Db  31551 AATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 31492

QY  227 TGCATGTGCGCATTCAGATTTGGGCAACATGAAACGGAATTTAAATTTTAACTTTAA 286
      || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  31491 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 31432

QY  287 AATAAATTTAAATTTTGAGTAAAT 309
      ||| ||||| || | | | | | | | | | | | | | | | | | | | | |
Db  31431 AATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 31409

RESULT 7
US-09-801-861-3/c
; Sequence 3, Application US/09801861
; Patent No. US20020119544A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001098

```


Query Match 4.3%; Score 47; DB 10; Length 419;
Best Local Similarity 46.1%; Pred. No. 2.3;
Matches 158; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

```
QY 11 TTTATATATTTTGTCAATATGATGATGAAAGTCCAAAAATTTTGTAAACGA 70
DB 413 TTTATATATATTTTAAATATATATATATATATATATATATATATATAT 354
QY 71 AACTCCAAATTTTATTTTATGCAACAGAAATTAACAGATGAAACATTTTGTG 130
DB 353 ATTTTATATATATTTTATATATTTTATATATATATATATATATATATAT 234
QY 131 TGAATGGAAGTATATATATATATATATATATATATATATATATATATAT 190
DB 293 TTTAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 234
QY 191 GCTCAAGTATGTTATCTAGTATGATATATATATATATATATATATATAT 230
DB 233 TTTATATATATATATATATATATATATATATATATATATATATATAT 174
QY 251 ACAACATGAAACGAAATTAATATATATATATATATATATATATATATAT 310
DB 173 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 114
QY 311 TGTTCCTGACTATGAGGGCAAAAAAGACATGCCAAA 353
DB 113 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 71
```

RESULT 11

US-09-938-842A-4756
Sequence 4756, Application US/0993842A
Patent No. US20020160378A1

GENERAL INFORMATION:

APPLICANT: Harper, Jeff

APPLICANT: Krieps, Joel

APPLICANT: Wang, Xun

APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

FILE REFERENCE: SCRIPI300-3

CURRENT APPLICATION NUMBER: US/09/938, 842A

CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227, 866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264, 647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300, 111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 4756

LENGTH: 1713

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-938-842A-4756

Query Match 4.2%; Score 46.2; DB 9; Length 1713;

Best Local Similarity 48.3%; Pred. No. 5.7; Length 1713;

Matches 129; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
QY 51 AAAAAAAAAATTTGTTACGAACTCCAAATTTTATTTTATGAAACAGAAATAC 110
DB 874 ATATATATTTTATATATTTTAAATCTAAATCTAAATCTAAATCTAAATCT 933
QY 111 GATGAAACATATTTTGTGTGGAAGGAGTATATATATATATATATATAT 170
DB 934 CATATATATTTTGTATATATGCTAAATCTAAATCTAAATCTAAATCTAA 993
QY 171 AATTAATATAGCTTATGCGGCTCAAGTATGTTATCTAGTATGATTAATTA 230
DB 994 AATTTTAAATTTATATATATTTTGTATATATATATATATATATATAT 1053
```

```
QY 231 TGTGCGATTCAGATTTGGGACAAATGAAACGAAATTAATATATATAT 290
DB 1054 GCTTAATATTAAGGAGATGATGATGATGATGATGATGATGATGAT 1113
QY 291 AATTAATTTGATGATATGATGATGATGATGATGATGATGATGATGAT 317
DB 1114 TATGAAAGCTCAATCACTTTGATTC 1140
```

RESULT 12

US-09-996-956-5/C
Sequence 5, Application US/09996956
Patent No. US20020155463A1

GENERAL INFORMATION:

APPLICANT: Origene Technologies, Inc

TITLE OF INVENTION: Prostate Polynucleotides and Uses

FILE REFERENCE: 9U 301 R1

CURRENT APPLICATION NUMBER: US/09/996, 956

CURRENT FILING DATE: 2001-11-30

PRIOR APPLICATION NUMBER: 60/250, 354

PRIOR FILING DATE: 2001-12-01

NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn version 3.1

SEQ ID NO 5

LENGTH: 32463

TYPE: DNA

ORGANISM: Homo sapiens

US-09-996-956-5

Query Match 4.2%; Score 46; DB 9; Length 32463;

Best Local Similarity 47.0%; Pred. No. 19; Length 32463;

Matches 142; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

```
QY 39 TAGAAGTTCACAAATTTTGTATACAGAACTCCAAATTTTATTTTATGGA 98
DB 4922 TATTAAGTATATATTTTAAATCTAAATCTGAAACAGAAATTTTCTAC 4863
QY 99 CAGAAATTAACAGATGAACTATTTTGTGGAATGGAAGTATATATATAT 158
DB 4862 ACACAAAGAAATTTGAAACAAATTAATGAGTACAAAGATATATATAT 4803
QY 159 CAATTTTAAATTTTATTAAGCTATACGCGCTCAAGTATGTTATGATG 218
DB 4802 AAATTTTCAAGTGTGCTTGGATATCTAGTACAGTGGAGGAATCTTT 4743
QY 219 ATTAATATGATGCTGCGATTCGATTTGGACAAATGAAACGAAATTA 278
DB 4742 TCCAGTTTACTGATTAATCTTATATTTGACACAGAAATGAAATCA 4683
QY 279 AACTTTAAATTAATTAATTAATTTGAGTAAATGTTTCTGATATGAG 338
DB 4682 TATGATTTGATGATTAATTAAGTTTACAAAGTATGATATATATTA 4623
QY 339 AA 340
DB 4622 GA 4621
```

RESULT 13

US-09-960-352-573/C
Sequence 573, Application US/09960352
Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Mengling

APPLICANT: Byate, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960, 352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:08:52 ; Search time 1436.66 Seconds

(without alignments)
12411.569 Million cell updates/sec

Title: US-09-502-426a-1_COPY_2102_3202

Perfect score: 1101

Sequence: 1 aatcacaataatataat.....gagagagaactagcttc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 / Gapext 1.0

Searched: 16154066 segs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	148	13.4	148	17	BH811044
2	145.6	13.2	473	17	AL768954 Arabidops
3	69.6	6.3	1101	17	CNS0039G
c 4	69.4	6.3	1201	17	CNS0167M
5	66.8	6.1	928	17	CNS00DKY
c 6	66.4	6.0	1201	17	CNS0165X

7	65.8	6.0	1101	17	CNS0021J	AL061936 Drosophill
8	65.2	5.9	1101	17	CNS0047E	AL078714 Drosophill
c 9	64.6	5.9	1101	17	CNS00DFL	AL069706 Drosophill
10	63.6	5.8	1146	17	CNS021G2	AL176843 Tetradon
c 11	63.4	5.8	1101	17	CNS0021J	AL061936 Drosophill
c 12	63.2	5.7	1101	17	CNS0080B	AL057419 Drosophill
c 13	62.8	5.7	1009	17	CNS00587	AL057153 Drosophill
14	62.6	5.7	1101	17	CNS00DFL	AL069706 Drosophill
15	62.4	5.7	855	17	CNS04P5D	AL030874 Tetradon
c 16	62	5.6	1029	17	CNS014GM	AL174271 Tetradon
c 17	61.8	5.6	1101	17	CNS003BD	AL064091 Drosophill
18	61.6	5.6	609	17	CNS025K2	AL182171 Tetradon
c 19	61.6	5.6	1101	17	CNS016L1	AL106896 Drosophill
c 20	61.6	5.6	1190	17	CNS020N7	AL120698 Tetradon
21	61.6	5.6	1200	17	CNS016CO	AL106578 Drosophill
22	61.4	5.6	1092	17	CNS020K7	AL175636 Tetradon
23	61	5.5	910	17	CNS018BP	AL142826 Anophel
24	61	5.5	1085	17	AG075009	AG075009 Pan trogl
25	60.8	5.5	1101	17	CNS0080B	AL057419 Drosophill
c 26	60.8	5.5	1101	17	CNS001FB	AL060732 Drosophill
c 27	60.6	5.5	878	17	CNS0187R	AL108993 Drosophill
c 28	60.6	5.5	1101	17	CNS0039G	AL063921 Drosophill
c 29	60.2	5.5	1201	17	CNS016C2	AL106556 Drosophill
30	60	5.4	945	17	CNS014DK	AL285149 Tetradon
c 31	60	5.4	1225	17	CNS0161D	AL106171 Drosophill
c 32	59.8	5.4	964	17	CNS016BD	AL441457 T7 end of
c 33	59.8	5.4	966	17	CNS0057C	AL061991 Drosophill
c 34	59.6	5.4	1092	17	CNS020K7	AL175636 Tetradon
c 35	59.6	5.4	1368	10	BE420618	BE420618 HMM000.60
c 36	59.2	5.4	524	17	CNS01U90	AL167541 Tetradon
c 37	59.2	5.4	843	17	CNS00CS1	AL059666 Drosophill
38	59	5.4	987	17	CNS014PQ	AL104456 Drosophill
39	59	5.4	984	17	CNS04NDJ	AL1298972 Tetradon
40	58.8	5.3	1101	17	CNS008X3	AL052544 Drosophill
41	58.6	5.3	1001	17	CNS0135H	AL105023 Drosophill
42	58.6	5.3	1101	17	CNS00807	AL069440 Drosophill
c 43	58.4	5.3	676	17	CNS003XU	AL065304 Drosophill
c 44	58.4	5.3	843	17	CNS00CS1	AL059666 Drosophill
45	58.4	5.3	996	17	CNS00DFH	AL071063 Drosophill

ALIGNMENTS

RESULT 1
BH811044/c 148 bp DNA linear GSS 02-MAY-2002
LOCUS SALK_057128 Arabidopsis thaliana tDNA insertion lines Arabidopsis
DEFINITION thaliana genomic clone SALK_057128, DNA sequence.
ACCESSION BH811044
VERSION BH811044.1 GI:20388862
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 148)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Predits,I., Shinn,P.,
Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
tDNA.

TITLE JOURNAL
COMMENT

Class: TDNA tagged.
Location/Qualifiers
1. 148
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db.xref="taxon:3702"
/clone="SALK_057128"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at <http://signal.salk.edu/tdna-protocols.html>"

BASE COUNT 46 a 28 c 14 g 60 t
ORIGIN

Query Match 13.4%; Score 148; DB 17; Length 148;
Best Local Similarity 100.0%; Pred. No. 6.6e-18;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 ATGGAACAAGAAATACAGATAGAAACTATTTGTTGGAATGAGATGATATATAC 152
DB 148 ATGGAACAAGAAATACAGATAGAAACTATTTGTTGGAATGAGATGATATATAC 89
OY 153 ATTAAGCAAAATTTTAAAAATATATACCTATACGCCCTCAAGATATATCTAGTA 212
DB 88 ATTAAGCAAAATTTTAAAAATATATATACCTATACGCCCTCAAGATATATCTAGTA 29
OY 213 GGTGTAATTAATGATGATGTCGAT 240
DB 28 GGTGTAATTAATGATGATGTCGAT 1

RESULT 2
AL768954 473 bp DNA linear GSS 18-JUN-2002
LOCUS
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-082A08-011867,
genomic survey sequence.
ACCESSION AL768954
VERSION AL768954.1 GI:21522073
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
Stilzhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Seidler, H.
and Weisshaar, B.
TITLE A pipeline for automated high-throughput generation of ESTs
(flanking sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
JOURNAL unpublished
AUTHORS 2
Rosso, M., Stilzhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
REFERENCE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
for flanking sequence tag based reverse genetics
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 473)
Li, Y., Rosso, M., Stilzhov, N. and Weisshaar, B.
COMMENT Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer
Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone T3A5. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
plant genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.
FEATURES
SOURCE Location/Qualifiers
1. 473

/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db.xref="taxon:3702"
/clone="GK-082A08-011867"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 168 a 58 c 70 g 146 t 31 others
ORIGIN

Query Match 13.2%; Score 145.6; DB 17; Length 473;
Best Local Similarity 87.0%; Pred. No. 1.3e-17;
Matches 160; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 1 AATCTACAATTTATTTATTTAGTCATATACATGATGAAAGTCCAAAAATTTT 60
DB 290 AATCTACAATTTATTTATTTAGTCATATACATGATGAAAGTCCAAAAATTTT 349
OY 61 TGTTAACAGAACTTCCAAATTTTTTTTTTTATGCAACAGAAATACAGATGAAAC 120
DB 350 TGTTAACAGAACTTCCAAATTTTTTTTTTTGCGGCGCAAAATTTACCGATGAAAC 409
OY 121 TATTTGTTGGAAGGAGTAGTATATACATTAAGCAAAATTTTAAAAATATATTA 180
DB 410 TTTTGGGGGGGGAAGGAGTAGTATATATATATATATATATATATATATATTA 469
OY 181 GCCT 184
DB 470 CCTT 473

RESULT 3
CNS00396 1101 bp DNA linear GSS 03-JUN-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence TETS end of BAC #
BACR08K10 of RRC1-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION CNS00396
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
Genoscope.
COMMENT Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutyo Osoegawa and
Aaron Mamoser in Pletier de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named Rpci-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

REFERENCE
AUTHORS
TITLE
JOURNAL

Ephydriidae; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

COMMENT

FEATURES
source
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29B23"
/clone_1lb="RPCI-98"
/note="end : 17"

BASE COUNT
ORIGIN
419 a 91 c 60 g 299 t 232 others

Query Match 5.9%; Score 64.6; DB 17; Length 1101;
Best Local Similarity 34.3%; Pred. No. 0.014;
Matches 139; Conservative 71; Mismatches 195; Indels 0; Gaps 0;

OY 566 TTTATTATTAAAAACACACACTTGATGACATATAAATTGATAACCTTTAGCANG 645
DB 915 TTTATTTTATTTATTTATTTTAAAAAAMWMAAATTTTATTTTATTTATTTATA 856
OY 646 GATTCGATGATCTATCTATGACTTTTCTTACACGCGTGAATGATTTACT 705
DB 855 TTTTATTTTATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 796
OY 706 ACTATTAGCAGACAAATGATTTATGATATATCCATTAATCCATGATTTATGAT 765
DB 795 ATTATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 736
OY 766 AATAGCTTTAACTATTTGACATCCGACGCTTCTGCACTTTGTTTAAATTTAG 825
DB 735 ATTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 676
OY 826 AGTTATATTAATTAAGTATTAAGAGCATACAGGACCAAAAGTATGACACGG 885
DB 675 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 616
OY 886 AGAAAACAAAGCCATGAGCTGTTAGTTAACTTAATTAAGAGATTTATTA 945
DB 615 TTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 556
OY 946 TTTTATGACGATTAACATTAATTTTCTGACTTCTTTAAA 990
DB 555 TTAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 511

RESULT 10
LOCUS CNS021G2 1146 bp DNA linear GSS 12-MAY-2000
DEFINITION Tetradon nigroviridis genome survey sequence 17 end of clone
225004 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL176843
VERSION AL176843.1 GI:7814900

KEYWORDS
SOURCE
ORGANISM

GSS; genome survey sequence.
Tetradon nigroviridis.
Tetradon nigroviridis.

REFERENCE

AUTHORS
TITLE
JOURNAL
AUTHORS

Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Filzmes, C., Wnicker, P., Brotlier, P., Quetler, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 1146)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Filzmes, C., Wnicker, P., Brotlier, P., Quetler, F.,
Saurin, W. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater puterfish Tetradon nigroviridis
Unpublished
3 (bases 1 to 1146)
Weissenbach, J.

COMMENT

JOURNAL
AUTHORS
TITLE
JOURNAL
AUTHORS

Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetradon>.

FEATURES
source
1. 1146
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="225004"
/clone_1lb="G"
/note="Genoscope sequence ID : C0AG225BH02LPI-end : 17"

BASE COUNT
ORIGIN
393 a 60 c 73 g 274 t 346 others

Query Match 5.8%; Score 63.6; DB 17; Length 1146;
Best Local Similarity 31.8%; Pred. No. 0.021;
Matches 150; Conservative 98; Mismatches 222; Indels 1; Gaps 1;

OY 534 ACTCTCGGATTCGCTGAGCATTTATCAAAAATTTATGACGACATGCTTTATTA 593
DB 591 ATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 650
OY 594 TTTAAACCTCACACTGATGATCAATAAATTTGATTAACACTTTACGATTCGTA 653
DB 651 KTTTAAAGKKKTK 709
OY 654 CGATCTCTAATGACTTTTCTACACGCGTGAATGATTAATGATTAATG 713
DB 710 AGKGRKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRAKRA 769
OY 714 CCAGACAAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 773
DB 770 KRRKKRAAARAKKBRKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGKRGK 829
OY 774 GTTAACTATTCAGATGCGACTTTCGCACTTTGTTTATTTAATTAAGCTTAAT 833
DB 830 ATTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 889
OY 834 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 893
DB 890 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 949
OY 894 AAGCATGACGCTGATGATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 953
DB 950 AAAATTTTTRRDKRTTTTTRTTTATTTTATTTTATTTTATTTTATTTTATTTA 1009
OY 954 ACGATGATTAACATTTATTTTCTGACTTCTTTAAACCCCTCTTTACAAA 1004

DB	1010	ATAAATAATAAAAATTGTTTGTGTTTGGTTTAAANAAAMTTTAAWMAA	1060
RESULT 11	CNS0021J	1101 bp	DNA
LOCUS	CNS0021J/c		Linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL061936		
VERSION	AL061936.1	GI:4940214	GSS.
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
AUTHORS	determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's PL and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
FEATURES	source	location/Qualifiers	
	1..1101	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"		
	/clone="BACR05N11"		
	/clone_1lb="RPCI-98"		
	/note="end : TET3"		
BASE COUNT	631 a	7 c	28 g 289 t 146 others
ORIGIN			
Query Match	5.8%	Score 63.4;	DB 17; Length 1101;
Best Local Similarity	49.2%;	Pred. No. 0.023;	
Matches	88;	Conservative 25;	Mismatches 66; Indels 0; Gaps 0;
OY	2 ATCTACAATTTTAATTTAGTCATATACATGCATGAGAAGTCCAAAAAATTTT	61	
Db	: : : : :		
742	TTTTTAAAAMTATATATATTTTAAAWMAAAAAAAAMTAAAAATTTTAAATTTTAA	683	
OY	62 GTTACAGAACTCCCAATTTTTTTTTTTATGGAACAAGAAATACAGATAGAAACT	121	
Db	: : : :		
662	ATWATATAAAMAATWTATWTATTTTWTATTTATTAATTAANAANAANAAMATTTTAA	623	
OY	122 ATTGTTGTGGAATGGAAGTAGATATACATTAAGCAAAATTTTAAAAAATATATATA	180	
Db	: : : :		
622	TTTTTWWTTMATTAATAAAMTAAAMAMTWTTTATTAATTTAAATTTTWWTTTAA	564	
RESULT 12	CNS00BOI/c	1101 bp	DNA
LOCUS	CNS00BOI/c		linear GSS 04-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC # BACR2AD09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL057419		
VERSION	AL057419.1	GI:4937885	

KEYWORDS	GSS.
SOURCE	Drosophila melanogaster.
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
JOURNAL	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and Esg1 libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .
FEATURES	Location/Qualifiers
Source	1..1101
	/organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACR24D09" /clone_1kb="RPCI-98" /note="End : 17"
BASE COUNT	420 a 56 c 48 g 261 t 316 others
ORIGIN	
Query Match	5.7%; Score 63.2; DB 17; Length 1101;
Best Local Similarity	41.1%; Pred. No. 0.025;
Matches 176; Conservative 51; Mismatches 199; Indels 2; Gaps 2.	
558	TTTATCAAAAATTTATTAGCAGCAATGGTTTATTATTTAAAAACTCACACTGTGATCAG 617
Db	1048 TTTWMTATTTTAAWMTATATATWATTTTAAAAAAATWAAWTTTATTATWMAAAT 989
Qy	618 ATAAAAATTCATAAACACTTTACGATGATCGACGATCTAATGACTTTT 677
Db	988 TTTAAATTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 929
Qy	678 TTCTACACAGCGTGAAAGTTATAGTCTATTAGCCGAGACCAATGATATGATAT 737
Db	928 WATTAATAAAAAATATA-AMAAATATTATTATTTTAAATTAATTTTATTTATTT 870
Qy	738 ATCCATTAATTCATGATATTTATGATATAAATAGCTGTTAACTATTTCAGATCGCAG 797
Db	869 WTTTATTATTWATTAATAATTTATAMAWMTATWTTTATTATTTTATTTTATTT 810
Qy	798 TTTCGCA-ACCTTTGTTTATTATTAGAGTTTAAATATAAGTATTAAGAGACA 856
Db	809 TTTTAAATWATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 750
Qy	857 TAACGAGGCAACAAAGTATGACACGAGAAACAAACCATGACGATCGATTGGTTAG 916
Db	749 AAAAAAAMATATAAAAAATTTTTTTTTTTTTTTTAAATTAATTAATTAATTAATTA 690
Qy	917 TTTAAGCTTAATTAAGACATTTTATTAATTTTAAATTAAGCATGATATACATTATTTTC 976
Db	688 AATAATWATAATAATSVACBSBSVTTTATATTTTAAATAAAMAKAMAAATTTTGTGSTR 630
Qy	977 TGACTTCT 984
Db	629 TTTATTTT 622

RESULT 13
CNS00587/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CNS00587 1009 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TE13 end of BAC #
BAC1101 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL057153.1 GI:4932541
GSS.
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1009)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1..1009
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BAC1101"
/clone_11b="RPCI-98"
/note="end : TE13"

BASE COUNT 263 a 116 c 96 g 477 t 57 others

ORIGIN

Query Match 5.7%; Score 62.8; DB 17; Length 1009;
Best Local Similarity 46.5%; Pred. No. 0.031;
Matches 148; Conservative 10; Mismatches 160; Indels 0; Gaps 0;

QY 1 AATTCACAAATTAATTTTATGTCATACATGATGAGAAAGTCCAAAAAATTTT 60
DB 1006 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 947
QY 61 TGTAAACAGAACTCCAAATTTTATGAGACAGAAATATACAGATAGAAAC 120
DB 946 NATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 887
QY 121 TATTTGTTGTAATGAGATGATATACATTAAGCAATTTTAAAAATATATATA 180
DB 886 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 827
QY 181 GCCTATAGCGGCTCAAGATGATCTATCTAGTGGTAAATTAATTAATGATGGTGGATT 240
DB 826 TAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 767
QY 241 CAGAAATTTGGACAAATGAGAAATTAATTAATTAATTAATTAATTAATTAAT 300
DB 766 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 707
QY 301 TGAGTAATGTGTTTCT 318
DB 706 TWAATTAATTAATTTTWT 689

RESULT 14
CNS00587/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CNS00587 1101 bp DNA linear GSS 04-JUN-1999
Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR29823 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL069706.1 GI:4949849
GSS.
Drosophila melanogaster.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29823"
/clone_11b="RPCI-98"
/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match 5.7%; Score 62.6; DB 17; Length 1101;
Best Local Similarity 35.1%; Pred. No. 0.032;
Matches 140; Conservative 65; Mismatches 194; Indels 0; Gaps 0;

QY 8 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 67
DB 587 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 646
QY 68 AGAACTCCAAATTTTATGAGACAGAAATTAATTAATTAATTAATTAATTA 127
DB 647 TATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 706
QY 128 TGTGATGAGAGATGATATATACATTAAGCAATTTTAAAAATTAATTAAGCCATA 187
DB 707 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 766
QY 188 CGGCTCAAGATGATCTATAGTGGTAAATTAATTAATTAATTAATTAATTA 247
DB 767 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 826
QY 248 GGCACAAATGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 307
DB 827 WATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 886
QY 308 ATGTGTTTCTGACTATTTAGGGGCAAAAAAGACAAAGTCCAAAGTCTACGGTTGA 367

9
1
1
1

10

11

12

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:06:10 ; Search time 31531.6 Seconds

(without alignments)
2955.362 Million cell updates/sec

Title: US-09-502-426a-1_COPY_1_3202

Perfect score: 3202

Sequence: 1 atgtggatattatattgtt.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 205640 seqs, 14551402878 residues 4109280

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : GenEmbl.*
1: gb_da.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hg_hum.*
31: em_hg_inv.*
32: em_hg_other.*
33: em_hg_mus.*
34: em_hg_pln.*
35: em_hg_rod.*
36: em_hg_mam.*
37: em_hg_vrt.*
38: em_sy.*
39: em_hggo_hum.*
40: em_hggo_mus.*
41: em_hggo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	3134.2	97.9	84196	8	ATT3A5	AL132979 Arabidops
C 2	1132	35.4	4818	8	AF044216	AF044216 Arabidops
C 3	176.8	5.5	191	11	AL844732	AL844732 Arabidops
C 4	133.2	4.2	104992	2	AC005504	AC005504 Plasmodu
C 5	133.2	4.2	169586	2	AC004157	AC004157 Plasmodu
C 6	130.8	4.1	104992	2	AC005504	AC005504 Plasmodu
C 7	130.8	4.1	169546	2	AC004157	AC004157 Plasmodu
C 8	129.4	4.0	14867	3	AE001398	AE001398 Plasmodu
C 9	128.6	4.0	172816	9	AC093899	AC093899 Homo sapl
C 10	127.8	4.0	86827	3	PFMAL385	AL034556 Plasmodu
C 11	126.4	3.9	133877	2	AC120883	AC120883 Homo sapl
C 12	125.4	3.9	86827	3	PFMAL385	AL034556 Plasmodu
C 13	124.6	3.9	205130	2	AC105425	AC105425 Homo sapl
C 14	124.4	3.9	4601	3	DM011584	U1584 Drosophila
C 15	124.4	3.9	19517	3	DM037541	U37541 Drosophila
C 16	123.4	3.9	67970	3	PFMAL383	AL031746 Plasmodu
C 17	122.6	3.8	4601	3	DM011584	U1584 Drosophila
C 18	122.6	3.8	19517	3	DM037541	U37541 Drosophila
C 19	121.2	3.8	177293	2	AC069525	AC069525 Homo sapl
C 20	121	3.8	172816	9	AC093899	AC093899 Homo sapl
C 21	121	3.8	178670	9	AC104073	AC104073 Homo sapl
C 22	118	3.7	67970	3	PFMAL383	AL031746 Plasmodu
C 23	118	3.7	84196	8	ATT3A5	AL132979 Arabidops
C 24	116.4	3.6	72243	9	AL731858	AL731858 Human DNA
C 25	115.6	3.6	205130	2	AC105425	AC105425 Homo sapl
C 26	115.2	3.6	133877	2	AC120883	AC120883 Homo sapl
C 27	114.4	3.6	178670	9	AC104073	AC104073 Homo sapl
C 28	113	3.5	14867	3	AE001398	AE001398 Plasmodu
C 29	112.6	3.5	137889	9	AC073269	AC073269 Homo sapl
C 30	112.2	3.5	206059	2	AC127383	AC127383 Homo sapl
C 31	111.6	3.5	104853	9	AC117444	AC117444 Homo sapl
C 32	111.4	3.5	111861	9	AC069435	AC069435 Homo sapl
C 33	110.2	3.4	15635	3	AB083339	AB083339 Bombyx mo
C 34	110.2	3.4	15635	3	AB070264	AB070264 Bombyx mo
C 35	109.6	3.4	159618	9	AC096750	AC096750 Homo sapl
C 36	107.8	3.4	152209	9	HS1108D11	AL034419 Human DNA
C 37	107	3.3	105238	9	AC011458	AC011458 Homo sapl
C 38	106.6	3.3	106958	9	AL807813	AL807813 Human DNA
C 39	106.4	3.3	155106	9	AC104069	AC104069 Homo sapl
C 40	106.2	3.3	181792	9	AC098822	AC098822 Homo sapl
C 41	105.6	3.3	176174	9	AC007483	AC007483 Homo sapl
C 42	105.6	3.3	177293	2	AC069525	AC069525 Homo sapl
C 43	105.2	3.3	206059	2	AC127383	AC127383 Homo sapl
C 44	104.8	3.3	114276	9	AC011355	AC011355 Homo sapl
C 45	104.6	3.3	127439	9	AC126178	AC126178 Homo sapl

ALIGNMENTS

RESULT 1
ATT3A5 84196 bp DNA linear PLN 26-JAN-2000
LOCUS Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
DEFINITION AL132979
ACCESSION AL132979
VERSION AL132979.2 GI:6782244
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 84196)
Bloeker, H., Mewes, H.W., Lemcke, K., Meyer, K.F.X., Quetler, F. and
Salanoubat, M.

Pred. No. is the number of results predicted by chance to have a

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 84196)
AUTHORS Eu Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
lencx@mps.biochem.mpg.de; mayer@mps.biochem.mpg.de; Project
Coordinator: Marcel Salanoubat and Francis Queller, Groupement
d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
Gaston Cremieux, BP191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
On Jan 27, 2000 this sequence version replaced gi:6434247.
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
Location/Qualifiers
1. 84196
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
2983. .3338
/note="355 bp LTR"
3339. .7564
/note="LTR-transposon"
7565. .7922
/note="367bp LTR"
12339. .13197
/gene="T3A5.10"
join(12339. .12665,12742. .12803,12888. .13055,13125. .13197)
/gene="T3A5.10"
/note="similarity to cyclin-dependent kinase inhibitor
(ICK1), Arabidopsis thaliana, EMBL:AT094772.1"
/codon_start=1
/product="putative protein"
/protein_id="CAB62432.1"
/db_xref="GI:6561966"
/translation="MAVRRRRERDVVEENGVTITTVYRRKKEEVDLYESRILISPCY
QATNRGIVANSGASETSVYVRRKDSPEVEQCIIEEDSSVSCSTSEESKRR
KEVPLEENNDDRETETSMYTDLNKSESMNDSVAVEDVESRRRLKSLHETV
KEVLEDFQVAEKDLNKKLECSMKYNFDEKDEPLGGRYEWVKLP"
12339. .12665
/gene="T3A5.10"
/number=1
12666. .12741
/gene="T3A5.10"
/number=1
12742. .12803
/gene="T3A5.10"
/number=2
12804. .12887
/gene="T3A5.10"
/number=2
12888. .13055
/gene="T3A5.10"
/number=3
13056. .13124
/gene="T3A5.10"
/number=3
13125. .13197
/gene="T3A5.10"
/number=4
13198. .16470
/gene="T3A5.20"
complement(join(15858. .15901,16014. .16470))
/note="similarity to predicted proteins, Arabidopsis
thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB62433.1"
/db_xref="GI:6561967"
/translation="MARLEHNSMNGCAARISFSENFVEIRSEKSNNAKSNINSSSPS
MPSADFAFSYTDYSMPADELFLKGLTPPKRSHVARTIGEELLTREEGSMYDGNF

SLRPILLSSSFSRTGRTWRELGLKRTVHRSKTKDKVNEVLSODHKITISGNVATREC
QVADPR"
complement(15858. .15901)
/number=1
complement(15902. .16013)
/number=1
complement(16014. .16470)
/number=2
18303. .19931
/gene="T3A5.30"
complement(18303. .19931)
/codon_start=1
/product="scarecrow-like 7 (SCL7)"
/protein_id="CAB62434.1"
/db_xref="GI:6561968"
/translation="MAVWCCTDSGNLMAIQAOLIKOKOOQSOHOOEOEOEPNPWP
PSFGFTLPSGSDPFOVYNTNGFHPFLLEHONAAVASEEFDDEWESLINDGSA
QPNDFPRTYGHDPVSPSPRSASYSILNRVKKDSASQQLPPRPASTRINSPSPSP
HPPPPOPDLDNPIKAIHDIARKKETPDLIRIKESVSESGDYIYGFYAE
ALSHKETESPSSSSSLDEPLISYKTLNDACPYSKFAHLANQAIILEATQSNHII
VDGIFQGIQWSALQALATRSRSGKPTIRISGIPASISGDSPGSLATGNRLDFA
AILDNFEFYPVLPIQLNNGSFRVDEVLVYVFMELYKLDDETATGTALRLA
RSLNRPVYTGEEVSLNRYEFANRVKNSLRFSAVFESELPNDIDROSKERLVRVLL
FGRIMDVRSDDNNKRGCTRGLEKEEKORVLMERAGFEPPVAFSNVAQAKLLLM
NNSTLSLSESEGEFTSLANNVPLTVSSMR"
complement(18303. .19931)
/number=1
26093. .28999
/gene="T3A5.40"
complement(join(26093. .26338,27091. .27300,27308. .27386,
27478. .27570,27723. .28037,28136. .28288,28374. .28698,
28779. .28999))
/note="Contains Cytochrome P450 cysteine heme-Iron ligand
signature AA455-464; Prokaryotic membrane lipoprotein lipid
attachment site AA320-330"
/codon_start=1
/product="steroid 22-alpha-hydroxylase (DWF4)"
/protein_id="CAB62435.1"
/db_xref="GI:6561969"
/translation="MFEETHHTLPLLLPELSLLFLILKRRNRKTRNLPFGKS
GMPLEIGLKYPTATTLGDFMOHVSRYKGYLRSNLFGPEPTIVSDAGLNRIILO
NEGRLPESYPRSTIGLIGKSMVLIVGDHMRDRSTISLNFSAARLPTILLKVERH
TLFVLDNCOONSIFSAODEAKKFTFNMAKIKINSMDPEEETDQLKKRYVFMKGVAS
APLVPGTATYKALQSRATILKTERKMEKKLIDKEEDQEEVEKTEDEAKMSDSH
VRKQRTDDDLGWLKHSNSTEGLDILSLSLPAGHETSSVAIALAIFFAQACPRAV
EELREHELEIARAKKEGSESLNDMDYKMDFTCCVINEITSLGNYVFFLRKALKAV
RYKGYDIPSGWKVLPVTSAYVLDNSRYDQPLFPMRMOQONNGSSGSGSFYWG
NYMPFGGGRPCAGSELAKLEMAVFIIHLVLKFWMLAEDKPPAPFPVDPNCLPIR
VSRIL"
complement(26093. .26338)
/number=1
complement(26339. .27090)
/number=1
complement(27091. .27200)
/number=2
complement(27201. .27307)
/number=2
complement(27308. .27386)
/number=3
complement(27387. .27477)
/number=3
complement(27478. .27570)
/number=4
complement(27571. .27722)
/number=4
complement(27723. .28037)
/number=5
complement(28038. .28135)
/number=5
complement(28136. .28288)
/number=6
complement(28289. .28373)
/number=6

```

exon      complement(28374..28698)
          /number=7
intron    complement(28699..28778)
          /number=7
exon      complement(28779..28999)
          /number=8
gene       38307..41323
          /gene="T3A5.50"
CDS        complement(40180..40242,40337..40401,40489..40564,40655..40734,
          40805..40919,41225..41323)
          /note="Contains Eukaryotic putative RNA-binding region
          RNP-1 signature AA179-186"
          /codon_start=1
          /product="U1 snRNP 70K protein"
          /protein_id="CA62436.1"
          /db_xref="GI:6561970"
          /translation="MDSGDPLRNPNAVOARAVONRANVQLKMGSHPTGLTN
          NLKLEPPPLEYKPPPEKRCSPYTGMAOVSNFAPGDEYAPRPPEVELSOKR
          ERIHKLREKGYEKAAEDLKKDPNNDPATGDPKTLFVSLNANESSESKIKREFS
          VPKIRVHYVTDQLTNKRGIAFIEYMTTROMKAKDQCKIDGRVLYDVERGRT
          KSRKLSHQPRERSRDRKREKHNHRRDQGRDRDRDRDRDRDRDRDRDRDRDR
          GNDRTSRDHRDRSRKREDEYEGEYEHGGGRSRERNAEYRGSEPEETRGTYEDQGD
          TDRYSHRYDKMEEDPFYREYERKRSRSESEHYA"
exon      complement(38307..38796)
          /number=1
intron    complement(38797..38876)
          /number=1
exon      complement(38877..39090)
          /number=2
intron    complement(39091..39188)
          /number=2
exon      complement(39189..39270)
          /number=3
intron    complement(39271..40179)
          /number=3
exon      complement(40180..40242)
          /number=4
intron    complement(40243..40336)
          /number=4

Query Match      97.9%; Score 3134.2; DB 8; Length 84196;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 3186; Conservative 0; Mismatches 13; Indels 4; Gaps 4;

QY 1 ATGTGGTATTATATTTGTGGGTCGGTTGAGCTACATATTAATTCGTTTCGGT 60
DB 32199 ATGTGGGATATATATTTGTGGGTCGGTTGAGCTACATATTAATTCGTTTCGGT 60
QY 61 TATTCGTTCACATGATTTGAGTTGTTCTCAATTTGATCCAGATTAATTAATTT 120
DB 32139 TATTCGTTCACATGATTTGAGTTGTTCTCAATTTGATCCAGATTAATTAATTT 120
QY 121 AAAATTCATTTAAATATTTACAAGTAATTAATTAATTCATTTGATTTGATTAACA 180
DB 32079 AAAATTCATTTAAATATTTACAAGTAATTAATTAATTCATTTGATTTGATTAACA 180
QY 181 AATATCATCTTGGTATAGAGAAATATGAGTTTGAATTTATATATTAATTAAGGAA 240
DB 32019 AATATCATCTTGGTATAGAGAAATATGAGTTTGAATTTATATATTAATTAAGGAA 240
QY 241 TAATCATCTTGGTATAGAGAAATATGAGTTTGAATTTATATATTAATTAAGGAA 300
DB 31959 TAATCATCTTGGTATAGAGAAATATGAGTTTGAATTTATATATTAATTAAGGAA 300
QY 301 AATTCAGTAAATCAAAAAGAGATTTGAAGTGAATCAATTTGATTTGATTTGATTT 31900
DB 31899 AATTCAGTAAATCAAAAAGAGATTTGAAGTGAATCAATTTGATTTGATTTGATTT 31900
QY 361 AAAAAAATTTAAATCAAAAAGAGATTTGAAGTGAATCAATTTGATTTGATTTGATTT 420
DB 31839 AAAAAAATTTAAATCAAAAAGAGATTTGAAGTGAATCAATTTGATTTGATTTGATTT 420

```

```

QY 421 AGGTTATTTGATCATATTTACATGACATATTTCTTGTATTTACTATGAGATTTCTTAC 480
DB 31780 AGGTTATTTGATCATATTTACATGACATATTTCTTGTATTTACTATGAGATTTCTTAC 480
QY 481 CAATTAATTTGATCATATTTACATGACATATTTCTTGTATTTACTATGAGATTTCTTAC 540
DB 31720 CAATTAATTTGATCATATTTACATGACATATTTCTTGTATTTACTATGAGATTTCTTAC 540
QY 541 GTTTATTCATCTCCATCATCATGATTTATGATTTCTTGTATTTACTATGAGATTTCTTAC 600
DB 31660 GTTTATTCATCTCCATCATCATGATTTATGATTTCTTGTATTTACTATGAGATTTCTTAC 600
QY 601 ACAGTATTTGAGAAAAACGAAATTAAGAAAGGAAAGGATTTGATTTGATTTGATTTGATTT 660
DB 31600 ACAGTATTTGAGAAAAACGAAATTAAGAAAGGAAAGGATTTGATTTGATTTGATTTGATTT 660
QY 661 TGAATTAATTTCAAAAGAGATTAAGAGATGACACCAAAAGGTTGATTTGATTTGATTTGATTT 720
DB 31540 TGAATTAATTTCAAAAGAGATTAAGAGATGACACCAAAAGGTTGATTTGATTTGATTTGATTT 720
QY 721 GCCAGCTTTCTCTGACATCATATTTGACACCTATTTGATTTCTGATTTGATTTGATTTGATTT 780
DB 31480 GCCAGCTTTCTCTGACATCATATTTGACACCTATTTGATTTCTGATTTGATTTGATTTGATTT 780
QY 781 TTTCGATTAACGATTTGAAAAATATTTTATTTGATTTGATTTGATTTGATTTGATTTGATTT 840
DB 31420 TTTCGATTAACGATTTGAAAAATATTTTATTTGATTTGATTTGATTTGATTTGATTTGATTT 840
QY 841 GGTATTTGATTAATCTTCTGTTTAAAGCATTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 900
DB 31360 GGTATTTGATTAATCTTCTGTTTAAAGCATTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 900
QY 901 TAACTATTAATTTATCGGGATATATGATTAACATGATATATGACAAATTTGCTGGG 960
DB 31300 TAACTATTAATTTATCGGGATATATGATTAACATGATATATGACAAATTTGCTGGG 960
QY 961 ACCATTTGATTAATCTTCTGTTTAAAGCATTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1020
DB 31240 ACCATTTGATTAATCTTCTGTTTAAAGCATTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1020
QY 1021 CGATTTTACAGCGCTGATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1080
DB 31181 CGATTTTACAGCGCTGATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1080
QY 1081 TATTTATCAAAAGTATCAAAAGTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1140
DB 31121 TATTTATCAAAAGTATCAAAAGTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1140
QY 1141 CCGGTTGTTATGTAATATAGATTTAGTGTAAACAAATGTTTAAATTTGGAATTTA 1200
DB 31061 CCGGTTGTTATGTAATATAGATTTAGTGTAAACAAATGTTTAAATTTGGAATTTA 1200
QY 1201 ACATTAATATATTTCTTACAGAAAAAAACCTTAAGAGATTTAATCATATTTAT 1260
DB 31001 ACATTAATATATTTCTTACAGAAAAAAACCTTAAGAGATTTAATCATATTTAT 1260
QY 1261 GGGTATGCTATACCTTTACGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1320
DB 30941 GGGTATGCTATACCTTTACGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1320
QY 1321 GATTAATTAATTTACACGCGGTGTAATTAATTTAGGACCGTATGATTTGATTTGATTTGATTT 1380
DB 30881 GATTAATTAATTTACACGCGGTGTAATTAATTTAGGACCGTATGATTTGATTTGATTTGATTT 1380
QY 1381 ATATCATCTTCTGTTGTCACAAATTAATAAAGGAAAGGATTTGATTTGATTTGATTTGATTT 1440
DB 30821 ATATCATCTTCTGTTGTCACAAATTAATAAAGGAAAGGATTTGATTTGATTTGATTTGATTT 1440
QY 1441 CTGATTTCAATTAATGATTTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1500
DB 30761 CTGATTTCAATTAATGATTTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1500

```

QY 1501 TCTACACCGGTGTAACCATCTGCACATTAATTCCTTCTTTAATGACATCTTAACAT 1560
 |||||
 Db 30701 TCTACACCGGTGTAACCATCTGCACATTAATTCCTTCTTTAATGACATCTTAACAT 30642
 QY 1561 ATTATTTGTAGTGAATTAATTAAGAGCAACTGTATACATTAATTAATTAATAG 1620
 |||||
 Db 30641 ATTATTTGTAGTGAATTAATTAAGAGCAACTGTATACATTAATTAATTAATAG 30582
 QY 1621 ATACTAGTATGTATTAATTCACAAATACATCTTGGATGTTAACTTAATCTTCTTCT 1680
 |||||
 Db 30581 ATACTAGTATGTATTAATTCACAAATACATCTTGGATGTTAACTTAATCTTCTTCT 30522
 QY 1681 TCCACCGGTAAATTAATTAATCAATGAGTAAATTAATTTGCTATTTTCGGAGATC 1740
 |||||
 Db 30521 TCCACCGGTAAATTAATTAATCAATGAGTAAATTAATTTGCTATTTTCGGAGATC 30463
 QY 1741 ATGAAGATTAACCTTAATGATCTTAATTTTGAATATGTAACCTTTTACATAGAT 1800
 |||||
 Db 30462 ATGAAGATTAACCTTAATGATCTTAATTTTGAATATGTAACCTTTTACATAGAT 30403
 QY 1801 AATTAACCGTATGTTTGTGCAATATGACAGCCTTACACCTGTGATGATCAATTTT 1860
 |||||
 Db 30402 AATTAACCGTATGTTTGTGCAATATGACAGCCTTACACCTGTGATGATCAATTTT 30343
 QY 1861 TCTGCAATTAATTAATTAAGCAATTCATGCTATCTATCAATGAGAAACAGCTGAT 1920
 |||||
 Db 30342 TCTGCAATTAATTAATTAAGCAATTCATGCTATCTATCAATGAGAAACAGCTGAT 30283
 QY 1921 ACATTTTAATTAATTAAGCAATTTTGAATATGTAATTTTACATTAATTAATTA 1980
 |||||
 Db 30282 ACATTTTAATTAATTAAGCAATTTTGAATATGTAATTTTACATTAATTAATTA 30223
 QY 1981 AATATGATGCTATTAATTAATTTCTATGTTCTTAATTAATTTTATTAATTAATTA 2040
 |||||
 Db 30222 AATATGATGCTATTAATTAATTTCTATGTTCTTAATTAATTTTATTAATTAATTA 30163
 QY 2041 TAAATACCTTAATTAATTAATGTTGTAATTCATTAATTCATTAATTAATTTT 2100
 |||||
 Db 30162 TAAATACCTTAATTAATTAATGTTGTAATTCATTAATTCATTAATTAATTTT 30103
 QY 2101 AATCTACCAATTAATTAATTAATTAATGTAATTAATTAATTAATTAATTAATTA 2160
 |||||
 Db 30102 AATCTACCAATTAATTAATTAATTAATGTAATTAATTAATTAATTAATTAATTA 30043
 QY 2161 TTGTTACGAACTTCCAAATTTTATTTTATGAAACAGAAATTAACATTAAGAAA 2220
 |||||
 Db 30042 TTGTTACGAACTTCCAAATTTTATTTTATGAAACAGAAATTAACATTAAGAAA 22983
 QY 2221 CTATTTGTGTTGGAATGGAATGATTAATTAATTAAGCAATTTTAAATTAATTA 2280
 |||||
 Db 22982 CTATTTGTGTTGGAATGGAATGATTAATTAATTAAGCAATTTTAAATTAATTA 22923
 QY 2281 AGCCTATACGCGCTCAAGATGTTATTAATTAATTAATTAATTAATTAATTAATTA 2340
 |||||
 Db 22922 AGCCTATACGCGCTCAAGATGTTATTAATTAATTAATTAATTAATTAATTAATTA 22983
 QY 2341 TCAGATTTGGGCAACAAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 2400
 |||||
 Db 22982 TCAGATTTGGGCAACAAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 22983
 QY 2401 TTGAGTAAATGTTTCTGCTATTTGAGGGGCAATTAATTAATTAATTAATTAATTA 2460
 |||||
 Db 22982 TTGAGTAAATGTTTCTGCTATTTGAGGGGCAATTAATTAATTAATTAATTAATTA 22943
 QY 2461 GGGTTTGAATGCTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
 |||||
 Db 22942 GGGTTTGAATGCTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 22983
 QY 2521 TAGGGTCTCTTCAATTTTCACTGTTTCACTGTTTCACTGTTTCACTGTTTCACTG 2580
 |||||
 Db 22982 TAGGGTCTCTTCAATTTTCACTGTTTCACTGTTTCACTGTTTCACTGTTTCACTG 22963
 QY 2581 TATCTTAAGGTATTTTGGAAATCCCAATTAATTAATTAATTAATTAATTAATTAAT 2640
 |||||

Db 22962 TATCTTAAGGTATTTTGGAAATCCCAATTAATTAATTAATTAATTAATTAATTAAT 22963
 |||||
 QY 2641 TGGGATTTGCTGAGCACTTTATCAAAAATTTATGACGAATGGGTTATTAATTAATTA 2700
 |||||
 Db 22962 TGGGATTTGCTGAGCACTTTATCAAAAATTTATGACGAATGGGTTATTAATTAATTA 229503
 QY 2701 AACTCACAATGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
 |||||
 Db 229502 AACTCACAATGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 229443
 QY 2761 ATCTATGATCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2820
 |||||
 Db 22942 ATCTATGATCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 229383
 QY 2821 ACAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2880
 |||||
 Db 229382 ACAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 229323
 QY 2881 CTATTTGAGCATCGAGCTTCTGCAACTTTTGTGTTTAAATTAATTAATTAATTAATTA 2940
 |||||
 Db 229322 CTATTTGAGCATCGAGCTTCTGCAACTTTTGTGTTTAAATTAATTAATTAATTAATTA 229263
 QY 2941 AAGTTTAAAGGAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3000
 |||||
 Db 22922 AAGTTTAAAGGAGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 229203
 QY 3001 TGAACCTGATGTTTAA-CTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTA 3059
 |||||
 Db 22922 TGAACCTGATGTTTAACTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTA 229143
 QY 3059 GATTAACATTAATTAATTTTCTGACTTTTAAACCCCTCTTAAACAGAAACCTCCCTT 29083
 |||||
 Db 229142 GATTAACATTAATTAATTTTCTGACTTTTAAACCCCTCTTAAACAGAAACCTCCCTT 229083
 QY 3120 TTGAGTAAAGTCCGATCCCAATCTTAAAGCAATTAATTAATTAATTAATTAATTAATTA 3179
 |||||
 Db 229082 TTGAGTAAAGTCCGATCCCAATCTTAAAGCAATTAATTAATTAATTAATTAATTAATTA 229023
 QY 3180 AGAGAGAGAGAACTAGCTCC 3202
 |||||
 Db 229022 AGAGAGAGAGAACTAGCTCC 229000
 |||||

RESULT 2
 AF044216 4818 bp DNA linear PLN 25-JUN-2001
 LOCUS
 DEFINITION Arabidopsis thaliana sterol 22-alpha-hydroxylase (DMF4) gene,
 complete cds.
 ACCESSION AF044216
 VERSION AF044216.1 GI:2935341
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana.
 Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 4818)
 Choe,S., Dikes,B.P., Fujioke,S., Takatsuno,S., Sakurai,A. and
 Feldmann,K.A.
 The DMF4 gene of Arabidopsis encodes a cytochrome P450 that
 mediates multiple 22alpha-hydroxylation steps in brassinosteroid
 biosynthesis
 JOURNAL Plant Cell 10 (2), 231-243 (1998)
 MEDLINE 98158690
 PUBMED 9490746
 REFERENCE
 2 (bases 1 to 4818)
 Choe,S., Dikes,B.P., Azpiroz,R. and Feldmann,K.A.
 Direct Submission
 Submitted (22-JUN-1998) Plant Sciences, University of Arizona,
 Tucson, AZ 85721, USA
 JOURNAL
 FEATURES
 location/Qualifiers
 1..4818
 source

gene
 /organism="Arabidopsis thaliana"
 /strain="Ws-2"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="8 cm from nga6"
 990..4398
 /gene="DMF4"
 /note="DMF4"
 /allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4"
 990..1055

mRNA
 TATA_signal
 /gene="DMF4"
 join(<1133..1353,1434..1758,1844..1996,2095..2409,
 2562..2654,2746..2824,2931..3040,3795..4398)
 /gene="DMF4"
 /product="steroid 22-alpha-hydroxylase"
 join(1133..1353,1434..1758,1844..1996,2095..2409,
 2562..2654,2746..2824,2931..3040,3795..4400)
 /function="prassinosteroid biosynthetic enzyme"
 /note="member of the cytochrome P450 superfamily; CYP90B1"
 /product="steroid 22-alpha-hydroxylase"
 /protein_id="AAC05093.1"
 /db_xref="GI:2935342"

CDS
 /translation="MFEHHHTLPLILPLSLILLILKRNKRTPNLPKGS
 GMPLEGTIGLKPYATATLGDPMOQHSKTKYRSTFEPTVSDAGLNPILQ
 NEGRLEFSYRSITSGADPEAKTFENLAKHINMSDPEEETDOLKEEYVFMKGS
 APNIPGTAIKHALQSRTILKTERKREERLIKEDQEEETKEDDEAKSSPH
 VRKQEDDLGLGWLKHSNSTEQILDLILSLPAGHETSSVALAIFPLACPKRV
 BELREHLEIARAKKEGESELMNDYKMDFTQCINETLGNVRFILKRY
 RYKQYDIPSGWKLPIVSAVLNDSRYDQDPLFPMHMOONNASSSGSSTWEN
 NYMPGGPRLAGSELAKLEMAVFHHVILKFWELAEEDDPFAFFVDPNGLPLR
 VSRIL"

BASE COUNT 1566 a 727 c 888 g 1637 t
 ORIGIN

Query Match 35.4%; Score 1132; DB 8; Length 4818;
 Best Local Similarity 100.0%; Pred. No. 1.1e-153;
 Matches 1132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2071 GAATCAATATATCCATTATATTTTGAATCTACAAATTTATATATTTAGTCAAT 2130
 DB 1 GAATCAATATATCCATTATATTTTGAATCTACAAATTTATATATTTAGTCAAT 60
 QY 2131 AACATGCAATGAAAGTCCAAAAAATTTGTTACAGAAAGTCCAAATTTT 2190
 DB 61 AACATGCAATGAAAGTCCAAAAAATTTGTTACAGAAAGTCCAAATTTT 120
 QY 2191 TTATGGAACAGAAATACAGATAGAAACTATTTGTTGTTGGAATGGAAGTAGATA 2250
 DB 121 TTATGGAACAGAAATACAGATAGAAACTATTTGTTGTTGGAATGGAAGTAGATA 180
 QY 2251 TACATTAAGCAAAATTTAAAAAATTTATATAGCCTATACGCGCTCAAGATGTATCTA 2310
 DB 181 TACATTAAGCAAAATTTAAAAAATTTATATAGCCTATACGCGCTCAAGATGTATCTA 240
 QY 2311 GTAGGTATATATATATGATGCGATTCAGAAATGGGACACAAATGAAAGCAAT 2370
 DB 241 GTAGGTATATATATATGATGCGATTCAGAAATGGGACACAAATGAAAGCAAT 300
 QY 2371 TAAAAATTAATTTAAAAAATTTATATAGTAAATGTGTTTTCGACTATTGAGG 2430
 DB 301 TAAAAATTAATTTAAAAAATTTATATAGTAAATGTGTTTTCGACTATTGAGG 360
 QY 2431 GCGAAAAAAGCAATGCCAAAGTCTACGGGTTGAGTGTCCACTGGTATATATCT 2490
 DB 361 GCGAAAAAAGCAATGCCAAAGTCTACGGGTTGAGTGTCCACTGGTATATATCT 420
 QY 2491 AATTAATCTGTCTTTGACGCGACGCGGTGAGGCTCTTCTGACATTTTACGTCT 2550
 DB 421 AATTAATCTGTCTTTGACGCGACGCGGTGAGGCTCTTCTGACATTTTACGTCT 480

QY 2551 ACCCTACTGCGAGGCCACCCCTTTCCATATCTTAAGGGTATTTGAAATCCCAT 2610
 DB 481 ACCCTACTGCGAGGCCACCCCTTTCCATATCTTAAGGGTATTTGAAATCCCAT 540
 QY 2611 TTAACCGATGAGACCCGACGACTCTCTGGATTCGCTGGAGCATTTATCAAAAT 2670
 DB 541 TTAACCGATGAGACCCGACGACTCTCTGGATTCGCTGGAGCATTTATCAAAAT 600
 QY 2671 TATTAGCAGAAATGGTTTATTTATTTAAAAACACAACTGATCAATATTTAT 2730
 DB 601 TATTAGCAGAAATGGTTTATTTATTTAAAAACACAACTGATCAATATTTAT 660
 QY 2731 AATCACTTTTACGATGATGCTGACGATCTATCAATGATCTTTTCTTACACAGCT 2790
 DB 661 AATCACTTTTACGATGATGCTGACGATCTATCAATGATCTTTTCTTACACAGCT 720
 QY 2791 GGATGAAAGTTATAGTACTATTAGCAGACAAATTTGATATATATCATTAATCC 2850
 DB 721 GGATGAAAGTTATAGTACTATTAGCAGACAAATTTGATATATATCATTAATCC 780
 QY 2851 ATGATTTTATGATATTAATATAGCTGTTAACTATTGACATCGACGCTTCTGCACT 2910
 DB 781 ATGATTTTATGATATTAATATAGCTGTTAACTATTGACATCGACGCTTCTGCACT 840
 QY 2911 TTGTTTTTAAATTTTAAAGTTTAAATTAATTAAGATTTAAAGAGCATTAACGAGCAACA 2970
 DB 841 TTGTTTTTAAATTTTAAAGTTTAAATTAATTAAGATTTAAAGAGCATTAACGAGCAACA 900
 QY 2971 AAGTATTAAGACAGGAGAAACAAAGCCATGAGTGCATGCTGTTATTTAAGCTTAATA 3030
 DB 901 AAGTATTAAGACAGGAGAAACAAAGCCATGAGTGCATGCTGTTATTTAAGCTTAATA 960
 QY 3031 AGAATTTTATTTAAATTTTAAATGACATGATTAACAATTTATTTCTGACTCTTAA 3090
 DB 961 AGAATTTTATTTAAATTTTAAATGACATGATTAACAATTTATTTCTGACTCTTAA 1020
 QY 3091 ACCCTCTTACAAACAGAGCTCCCTTTTCAAGTAAAGTCGATTCGCAATCTTAAG 3150
 DB 1021 ACCCTCTTACAAACAGAGCTCCCTTTTCAAGTAAAGTCGATTCGCAATCTTAAG 3210
 QY 3151 ACAAGCCATTACAAAGAGAAAGTGAAGTGAAGAGAGAGAAATAGCTCC 3202
 DB 1081 ACAAGCCATTACAAAGAGAAAGTGAAGTGAAGAGAGAGAAATAGCTCC 1132

RESULT 3
 AL844732
 LOCUS
 DEFINITION
 Arabidopsis thaliana transposon insertion STS SM.3.29533, sequence tagged site.
 AL844732
 ACCESSION
 AL844732.1 GI:21998461
 VERSION
 STS; STS, sequence tagged site.
 KEYWORDS
 Arabidopsis thaliana
 SOURCE
 Arabidopsis thaliana
 ORGANISM
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1
 Clarke,J.H., Bowles,B., Carter,J., Hart,D., McCullagh,B., Murphy,G., Langham,S., Leclerc,C., Jones,J.D.G. and Bevan,M.
 Unpublished
 2 (bases 1 to 191)
 REFERENCE
 AUTHORS
 Clarke,J.H.
 TITLE
 Direct Submision
 JOURNAL
 Submitted (25-JUL-2002) Clarke J.H., John Innes Centre, Colney Lane, Norwich, NR4 7UJ, UK
 COMMENT
 AT denotes an activation tag dissociation transposon within a single line, ET an enhancer trap dissociation transposon, GT a gene trap dissociation transposon, MT a mis-expression enhancer trap dissociation transposon, SM a defective suppressor mutant transposon. _3 denotes a sequence derived from the 3' end of the transposon, _5 denotes a sequence derived from the 5' end of the


```

OY 1065 AAGGACCCGTCAGCAATTTATACAAAGTACAAAGTATAGCTGGAATCCTTAA 1124
    || || || || || || || || || || || || || || || || || || || ||
Db 73404 AATATATATATATATATATATATATATATATATATATATATATATATATAT 73463
OY 1125 GAAATTTTGGATTTACCGGTTGTTATGTAATATACATTTAGTGTAAACAAATATGT 1184
    || || || || || || || || || || || || || || || || || || || ||
Db 73464 AATAATTAATATATATATATATATATATATATATATATATATATATATATAT 73523
OY 1185 AATCAATAGTGTCAACATATACATATCTCTACGAAAAACAACTTAAGACAGT 1244
    || || || || || || || || || || || || || || || || || || || ||
Db 73524 AATTAATAAACAATATATATATATATATATATATATATATATATATATATAT 73583
OY 1245 TAACATATCAATATATGCGATGCTATACCT-----TTCAGTATGCTATAGTA 1295
    || || || || || || || || || || || || || || || || || || || ||
Db 73584 TATATATATATATATATATATATATATATATATATATATATATATATATAT 73643
OY 1296 GACTAAAGAAATAGTATGATGTCGATTAATGAATTCACACGCGTGGTAATATATG 1355
    || || || || || || || || || || || || || || || || || || || ||
Db 73644 AATTAATAAATACATATATATATATATATATATATATATATATATATATATAT 73703
OY 1356 GGACCGTATGTCAGTACGATGCAATATCAATCTGTTGGTCAACATATATATAT 1415
    || || || || || || || || || || || || || || || || || || || ||
Db 73704 TTTAT-TAATTTAAGATATATATATATATATATATATATATATATATATAT 73762
OY 1416 ACAAGAAAAAAGAAACGATTTTCTGATTCATTCATGATCTAATATGCAATAGT 1475
    || || || || || || || || || || || || || || || || || || || ||
Db 73763 TATTTATATTTATACATATATATATATATATATATATATATATATATATAT 73822
OY 1476 CTTTGGGTACAGTTTCAAGTCTCTACAGCGTGAACATGCAATATATATAT 1535
    || || || || || || || || || || || || || || || || || || || ||
Db 73823 AATGGTTATTTTATTTTATATATATATATATATATATATATATATATATAT 1595
OY 1536 GCTTTCTTTAATGCAATCTTTAATATATATATGTTAGTGTGAATTTAAGACGCACT 1595
    || || || || || || || || || || || || || || || || || || || ||
Db 73883 TATATATATATATATATATATATATATATATATATATATATATATATATAT 73942
OY 1596 TCTAACAATTA-----CAATATATATATGATCTAGTATGATATATCCAAAT 1645
    || || || || || || || || || || || || || || || || || || || ||
Db 73943 TATTAATAATTAATTTATATATATATATATATATATATATATATATATATAT 74002
OY 1646 ACATATCTTGTGATTTTAACTTATCTGTTCTCTCTACGATATATATATATATAT 1705
    || || || || || || || || || || || || || || || || || || || ||
Db 74003 AATATATATATATATATATATATATATATATATATATATATATATATATAT 74062
OY 1706 GAGTAAAAA--AAGTTTGTCTTATTTTTCGATGCAATGAGATTAACCTATAGCT 1763
    || || || || || || || || || || || || || || || || || || || ||
Db 74063 AATTAATATTTTAAATATATATATATATATATATATATATATATATATATAT 74122
OY 1764 TAAATTTTGGAAATGTAACCTTTTACGATGATTAATTAACCGTATGTTTGTGCGC 1823
    || || || || || || || || || || || || || || || || || || || ||
Db 74123 TAAATTTATTAATATATATATATATATATATATATATATATATATATATATAT 74182
OY 1824 ATAAATGACGCTCTACACGATGATGCAATTT--TTTCGAAATTTTAAATTAG 1880
    || || || || || || || || || || || || || || || || || || || ||
Db 74183 AATAATATTTAATATATATATATATATATATATATATATATATATATATATAT 74242
OY 1881 AATTCATGCTACATCAATATAGAAACGCTGAGTATTTACATTTTAAATTTAAACACA 1940
    || || || || || || || || || || || || || || || || || || || ||
Db 74243 ATTAATATATATATATATATATATATATATATATATATATATATATATATAT 74302
OY 1941 AATTTTGGAAAAATGTTAATTTCTAACAAATTTATTAATAAATATGATGCTATATGTA 2000
    || || || || || || || || || || || || || || || || || || || ||
Db 74303 TTTATATATATATATATATATATATATATATATATATATATATATATATAT 74362
OY 2001 TTTCCATATGTTCTTAAATATTTTATATATAGTATATATATATATATATATATAT 2060
    || || || || || || || || || || || || || || || || || || || ||
Db 74363 TTTATATATATATATATATATATATATATATATATATATATATATATATAT 74422
OY 2061 AATAGTGTGCAATTCAAATATCTCATTAAT-----ATTTTGAATCTACAAATTAAT 2116
    || || || || || || || || || || || || || || || || || || || ||
Db 74423 TATTAATTAATATATATATATATATATATATATATATATATATATATATATAT 74482

```

```

OY 2117 AATATTAATCAATATACAAATGATGAAGTCCAAAAAATTTTGTATACAGAACT 2176
    || || || || || || || || || || || || || || || || || || || ||
Db 74483 TATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 74542
OY 2177 CCAAATTTTATTTTATATGACAGAAATATACAGATTAAGAAACTATTTGTGTGGA 2236
    || || || || || || || || || || || || || || || || || || || ||
Db 74543 ATTAATATATATATATATATATATATATATATATATATATATATATATAT 74602
OY 2237 TCGAAGTATATATATATATATATATATATATATATATATATATATATATAT 2296
    || || || || || || || || || || || || || || || || || || || ||
Db 74603 TATATATATATATATATATATATATATATATATATATATATATATATATAT 74662
OY 2297 AAGTATGTTACTAGTATGTTATATATATATATATATATATATATATATATATAT 2356
    || || || || || || || || || || || || || || || || || || || ||
Db 74663 ATTTAATTTAATTAATATATATATATATATATATATATATATATATATATAT 74720
OY 2357 AATGAAAAAGGAAATTAATATATATATATATATATATATATATATATATATAT 2404
    || || || || || || || || || || || || || || || || || || || ||
Db 74721 ATTTAATTTAATATATATATATATATATATATATATATATATATATATATAT 74768

```

```

RESULT 5
AC004157
LOCUS
DEFINITION
AC004157 169546 bp DNA linear HTG 12-AUG-2000
PROGRESS ***, 2 unordered pieces.
AC004157.8 GI:9797712
VERSION
HTG: HTGS_PHASE1.
KEYWORDS
SOURCE
ORGANISM
Plasmodium falciparum.
Plasmodium falciparum.
Eukaryota; Alveolata; Apicomplexa; Hemosporidia; Plasmodium.
1 (bases 1 to 169546)
Hymen, R.W., Fung, E.L., Qin, F., Rowley, D., Mao, J., Tamaki, T.,
Kurd, O.B., Conway, A.B. and Davis, R.W.
Plasmodium falciparum 3D7 chromosome 12
Unpublished
2 (bases 1 to 169546)
Hymen, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.
Direct Substitution
Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology
Center, Stanford University, 855 California Avenue, Palo Alto, CA
94304, USA

```

```

COMMENT
On Aug 12, 2000 this sequence version replaced gi:8810447.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 23466: contig of 23466 bp in length
* 23467 23666: gap of unknown length
* 23667 169546: contig of 145880 bp in length.
Location/Qualifiers
1. 169546
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"
/clone="PRYAC293"
/clone="3D7"
BASE COUNT 69871 a 15381 c 15705 g 68389 t 200 others
ORIGIN

```

```

Query Match 4.2%; Score 133.2; DB 2; Length 169546;
Best Local Similarity 44.7%; Pred. No. 1.5e-10;
Matches 1041; Conservative 0; Mismatches 1248; Indels 39; Gaps 12;
OY 108 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 166
    || || || || || || || || || || || || || || || || || || || ||
Db 84033 ATTAAGCTATACAAATATATATATATATATATATATATATATATATATATATATAT 84092

```


OY	1597	GTAACATTGACAATTTTATTTAGATACCTGATGCATTTCCAAATAACACTCTGG	1656
Db	8645	ATTATGATATATATATATATATATTAATAATTAATTAATGATATATAAATAATTAACATAT	8704
OY	1657	ANGTTTAACTTAACCTTGTCCTCCACGGTAATAATTAATCATCGAGGTAAAAA	1716
Db	8705	ACATATATATTTACCAATATATTATTATTATTTAATTAATTTTACTATTTTTAAAT	8764
OY	1717	AATTTTGCTATATTTGCCGATCGATCGAAGATAAACCTAATGACTTAATTTTTGAAA	1776
Db	8765	TAAATATTTATTTTAAATATATTTATTTATTTATTTAAATTAATTTAATAATATAT	8824
OY	1777	ATGTAAACCCTTTCCTCAGATTAATTAACCGATCTGTTTGTGGCAATAAGCACGCT	1836
Db	8825	TGATTTTATTTATTTATTTATTT -TTATTTTATTTATTTATTTATTTATTTATTTT	8883
OY	1837	CACACACGTCATATGCAATTTTTTCTGCMAATATTTAAATTAAGAATTCATATCTAT	1896
Db	8884	AATTTATTTAATTTATTTAAATTAATTTATTTAATTAATTAATTTATTTATTTATTTAT	8943
OY	1897	CAATGAGAACAGACGAGATTAATTAATTTAAAGACAAAATTTTGAAAATGT	1956
Db	8944	TTATTTAAATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAA	9003
OY	1957	TATATTTCTACAAATATTTAATAAATATAGTCCATATATGTTCTGTAA	2016
Db	9004	TATATTAATTTATTTATTTATTTATTTATTTAATTTATTTAATTTATTTAATATTA	9063
OY	2017	AATATTTTTTTTTTATTTAGTTATATAARCATATGAACCAATATAGTTGG-----T	2070
Db	9064	ATTTATTTTATATTTATTTATTTATTTATTTAATTTATTTATTTATTTATTTAATTAAT	9123
OY	2071	GAATCCAATATCTCCATTAATTTTGTGAATCTACAAATTAATTAATTTGTGCAAT	2130
Db	9124	TATTTATTTAATATATTTATTTATTTATTTAATTTATTTATTTATTTATTTATTTAT	9183
OY	2131	AACATGCAATGAAAGTCCAAAAAAATTTGTTACAGAAACCTGCCAAATTTTTTTT	2190
Db	9184	TATTTTGTTATTTATTTATTTTAAATTAATTAATTTATTTATTTATTTATTTATTTAT	9243
OY	2191	TTTATGACAAATAACAGATAGAAAAATATTTGTGTGGAAGAAGTAGTATA	2250
Db	9244	TTTAAATTTATTTATTTAATTTAATTTTAAATTTAATTTAATTTAATTTAATTTATTT	9303
OY	2251	TACATTAAGCAAATTTTAAAAAATTAATATACCTATACGGCTCAAGATATGTATCTCA	2310
Db	9304	TAAATTAATTAATTTATTTATTTTAAATTAATTAATTTTAAATTAATTTAATTTAATA	9363
OY	2311	GTAGGTATATTAATATGCAATGAGCGCATTCGAATTTGGACAAATGAAAAAGGAAT	2370
Db	9364	TTTATTTAAAAAATATTTTTTAAATTTATTT--TTATTTAATCTATATTTTATTTATAC	9421
OY	2371	TAAATATTTAATTTAAAAATTAATAAATTTGAGTAAATGCTTTCTGCATATGAGG	2430
Db	9422	ACCAATTTAATTTTAAAAAAATATATATCTTTTAAATATATATTTAAATATTAAT	9481
OY	2431	GGCAAAAAAAAAAGCAATGCCAAATCTCA	2459
Db	9482	ACTTAAATTTAAATATATATATATGATA	9510

RESULT	9
AC093899	
LOCUS	AC093899
DEFINITION	Homo sapiens BAC clone RP11-724016 from 2, complete sequence.
ACCESSION	AC093899 AC068884
KEYWORDS	
SOURCE	
ORGANISM	Homo sapiens
REFERENCE	
COMMENTARY	
FEATURES	
ORIGIN	
BASE COUNT	
REMARKS	

SOURCE ORGANISM

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE	1 (bases 1 to 172816)
AUTHORS	Sulston,J.E. and Waterston,R.
TITLE	Toward a complete human genome sequence
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE	99063792
PUBMED	9847074
REFERENCE	2 (bases 1 to 172816)
AUTHORS	Pearman,C., Haakenson,W. and Boyer,E.
TITLE	The sequence of Homo sapiens BAC clone RP11-72A016
JOURNAL	Unpublished (2001)
REFERENCE	3 (bases 1 to 172816)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (10-SEP-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	4 (bases 1 to 172816)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (05-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	5 (bases 1 to 172816)
AUTHORS	Waterston,R.
TITLE	Direct Submission
JOURNAL	Submitted (12-JUN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT	On Feb 5, '2002 this sequence version replaced at:15625013

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watsn.wustl.edu
----- Summary Statistics -----
Center project name: H_NHO724016
Drafting Center: WtBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The Rpci-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tanno, M., Catanesu, J.J. and de Jong, P.J. (1996) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 31:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
PICTOR: pbace3.6

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RP11-724016;
actual end is at base position 172816 of RP11-724016.

Unresolved simple sequence repeat from base position 22700 to 25900.

OY	398	GATGATACCTTTGGTGTCTGCAGTTAAGGTTATTGTCATCATATTACATGCATCTTCCCTT	4457
Db	23338	AATAATATATTTATATTTAAATATATATATATATATATATATTAACAACATATATTTATATATATAT	2329
OY	458	GATTACTCAATACAGATTTCTCTTCCCAATTAATAATTTGCA-----ATTCAATATC	506
Db	23298	TATATATTAACATATATTTATATTATATATATATATATATATTAACAATATATATATATTA	23357
OY	507	TTGATTTATTAATTAATAATACGAGTGCGAATATGCCGTTATCGATCACCTCCATCAGATTT	566
Db	23358	ATATATATTTATATTTATATATATATATATATATTAACAATATATTTATATTTATATATATTA	23411
OY	567	ATGATCTTGTGCTAATCCAGCAATATATTAACAAGAGTTTAGAAAAACCGAAAATA	626
Db	23418	TATTAATATATATATATATATATATTAACAATATATTTATATTTATATATATATATTAATAA	23477
OY	627	AGAAAAAGGAAGAAGTAGTGACCATGAGATATGCAATATATATCAAGACAGATPAAGA	686
Db	23478	TATATTTATTTTAAATATATATATATTTAAATATATATATATATATATATATATATATAT	23533
OY	687	ATGACACCAAAAAGTTGGAGATATATGGTCCCTGCCAGCTTCTCTC-----ACAA	738
Db	23538	TAAATATATTTATATATTTAAATATATTTATATTAATATATATTTATATTAATATATAT	23597
OY	739	TCAATATCGACCTATTTGGATTTTCTGATATTCGTATAAAAATTTGCAATACGATGTG	798
Db	23598	TAAATATTTAAATATATTTATATATTAATATATATATATATATATATATATATATATAT	23655
OY	799	AAAAATATTT-TATTTGTAGCTGATCTCAATATTTATGTCCAAGTATTTGCATATCTT	857
Db	23658	ATATATATTTATATTTAAATATATATACATATATATTTATATATATATATATATATAT	23717
OY	858	CAGTTTAAAGCATATTTCTCTCTTCTTTCTTTGCTTTGCTCTCTPACATATATAT---	914
Db	23718	ATATTTAT	23777
OY	915	CGGGATATATGATTAACATGATATATATCAAAAACAATGTCTGGACCATTTTGAATAA	974
Db	23778	TATTAATATATATTTATATATTTATATATATATATATATATATATATATATATATATAT	23837
OY	975	ACTTTTTCACAAACATTACGGGACACTGACCTGCACCCCTTAATATACGATTTTTACAGGT	1034
Db	23838	TTTATATTTAAATATATTAACATATATTTATATATATTTAAATATATATATATATATTA	23897
OY	1035	CACATGACTGAGTTTCTACGATTAAGACATTAAGAACCCGTTCAACGCTATTTATACAAAGT	1094
Db	23898	ATATATATATATATATTTATATATTTATATATATATATATATATATATATATATATAT	23957
OY	1095	TACAAACCTGAATATATAGCTGAAATCCCTTAGAAATTTTGGAAATTACCGGTGTATGTA	1154
Db	23958	ATATTTATATATATAGATATATTTATTTTAAATATATTTTAGTGATTAATATATATATAT	24017
OY	1155	AATATATGATTTAGTGTAACCAATATGTCTATCAATATAGTGTCACATATATACATAT	1214
Db	24018	TA-----ATATATTTATATATTAATAATATATATATATATATATATATATATATAT	24072
OY	1215	CCTTACAGAAAAACAACCTTAAGAGAGTTAAACATATCCATATATGGGTATGCTATACC	1274
Db	24073	ATATTTTATATTTATATTTTATATATATTTATATATATATATATATATATATATATATTT	24132
OY	1275	TTTACAGTATCTATACAGACTPAAGAAATAGTATATGATGTCGATTAATGAAATVC	1334
Db	24133	TTATTTATTTAAATTAAT	24192
OY	1335	ACAGCCGGTGAATATATATNTGGGACCGATATTTACGATCACTGCATAATCATTTCTTGGT	1394
Db	24193	AGTTATATATATATATATATATATTTTATACAGTTA-TATTAATATATATTAATTTTATAGT	24251
OY	1395	TGCTCAACATTAATAAACCAAAAACAAGAAAAACAACAAATATTTCTGTGGATTCGATTC	1454
Db	24252	TATATTAATATATATATATATTTTATACAGGTATATATATATATATATATATATATATAT	24311

QY	1455	AAAGACGTAAAAAGCATAGATCTTTGGGTTACAGTTGCAAGTCCTCAACAGCGCTA	1514
Db	24312	TATTAATATATTTATATATTTATATACAGTTATATAATATATTTATTTATACAGTTATA	24371
QY	1515	ACCAATCGACATATTAATAATGGCTTCTTTAATGCACTCTTAAACATATTAATGGTGGT	1574
Db	24372	TAAATATATTTATATTTTATACAGTTATATAA -ATATTTTATATTTTATACAGTTATAT	24430
QY	1575	GGAAATTAATAAAGCAACTTGTACATTTACAAATTTATATTTAGACTAGTAGTGA	1634
Db	24431	AAATATATTTATATTTTATACAGTTATATAATATATTTTATATTTTATACAGTTATATAA	24490
QY	1635	TTATTCGCAATACATACCTTTGGATGGTTAACTTAATCTGTGTTCTCCACGATATAA	1694
Db	24491	ATA-TATTTATATTTATACAGTTATATAATATATTTTATAGTTTATATACATTTATATAA	24549
QY	1695	TATTAATCATCGAGTAAAAAAGTTTGCTTATTTTGGCAGTCGATGAGATTAAC	1754
Db	24550	TATATTTATATTTTATACATTTGTATTTAATATATTTT--ATATATAATATATTTTA	24606
QY	1755	TATGACTTAAATTTTTTGGAAATGTACCCCTTTTCTACATAGATTAATACCGATATGT	1814
Db	24607	TATTTATATATTTATATATATAATATAATATATGATATTTATATAATATTTATATATAATA	24666
QY	1815	TTTGTGCCATATGACAGCCCTACAGCTGATAGTCAATTTTCTGCAAAATTTAA	1874
Db	24667	TATATTTGATTTTAAATATGTTTATATATATAATATATATTTATATTTAAATATATGTTTTA	24726
QY	1875	ATTAGCAATTCAACTACTATCATATAGAAAGAAACAGCTAGTATACATTTTAAATTTAA	1934
Db	24727	TATATCAATATATATTTATATTTTAAATATATATTTTACATATAATATATTTATATTTGA	24786
QY	1935	AGACAAATATTTTGGAAAAATGTATATTTCTTACAAATATTTTAAATATATGCGCTAT	1994
Db	24787	TATATATTTATATTTGATATATATATTTTAAATATATTAATTAATTAATACATTTGATATATAT	24846
QY	1995	AATGTATTTCCATGTTCTTAAATATTTTTTTTTTATTTAGTTATTAATCATATATGA	2054
Db	24847	TTTATATATATTAATAATATTTTACATTTGCATATATATTTTATATATATTAATATATTTTACA	24906
QY	2055	ACCAATAA-TATAGTGGTGAATCCAAATATCTCCATTAATATTTTGGAAATCTACAAAT	2113
Db	24907	TTTGTATATATTTATATATATATTTTATATATTTTATATTTTGAATATATGTTTATAGATTAAT	24966
QY	2114	ATTATATTTTACGATATACATGATGATGAAGTATCCAAAAAATTTTGTATACAGAAA	2173
Db	24967	ATATATTTTACGTTTATATATATTTTATATGATTAATAATATATTTTACGTTTAAATATATTTT	25026
QY	2174	CTTCCAAATTTTTTTTTTATGACAACAAGAAATATACAGATGAAGAAACATTTTGTGTG	2233
Db	25037	ATCTATTAATATATATTTATGTTTAAATATATTTTATGATTAATAATATATTTTACGTTTA	25086
QY	2234	GAATGAGATGATATATACATTAAGCAATTTTAAAAAT 2274	
Db	25087	ATATATATTTTATAGTAAATATATATTTTACGTTTATATAT 25127	

RESULT 10					
PPMAL3P5					
LOCUS	PPMAL3P5	86827 bp	DNA	linear	INV 04-MAY-2000
DEFINITION	Plasmodium falciparum MAL3P5 complete sequence.				
ACCESSION	AL034556 AL000891 AL008972 AL008978 AL010141 AL010153 AL010162				
VERSION	AL010206 AL010210 AL139179				
KEYWORDS	AL034556.3 GI:7711064				
SOURCE	HTG; centromere; CIMP protein; initiation factor E4; Serine/threonine protein phosphatase.				
ORGANISM	Plasmodium falciparum 3D7.				
REFERENCE	Plasmodium falciparum 3D7.				
AUTHORS	Enkaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. 1 (bases 1 to 86827)				
	Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T., Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T., Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holroyd,S.,				

/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 5 of
PFC0581w"
8369..8374
/gene="PFC0581w"
/note="potential splice donor sequence for exon 5 of
PFC0581w"
8559..8567
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 6 of
PFC0581w"
8642..8647
/gene="PFC0581w"
/note="potential splice donor sequence for exon 6 of
PFC0581w"
8704..8712
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 7 of
PFC0581w"
8813..8818
/gene="PFC0581w"
/note="potential splice donor sequence for exon 7 of
PFC0581w"
8913..8923
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 8 of
PFC0581w"
9007..9012
/gene="PFC0581w"
/note="potential splice donor sequence for exon 8 of
PFC0581w"
9112..9121
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 9 of
PFC0581w"
9239..9264
/gene="PFC0581w"
/note="potential splice donor sequence for exon 9 of
PFC0581w"
9361..9368
misc_feature

Query Match 4.08; Score 127.8; DB 3; Length 86827;
Best Local Similarity 44.08; Pred. No. 1.1e-09;
Matches 1051; Conservative 0; Mismatches 1307; Indels 29; Gaps 11;

QY 35 TACAATATAATTCGTTCTGTTCTGTTATCTGTTCAAGATGTTGTTGTTCTCAA 94
DB 38839 TAATATCAATATAATTCGTTCTGTTATCTGTTCAAGATGTTGTTGTTCTCAA 94
QY 95 TTGGATTCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 154
DB 38899 TATAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 154
QY 155 ATCTTACATGTTATGTTATTAACAAATATCTGTTGTTATGAGAAATATGAG 214
DB 38959 ATACATATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 214
QY 215 TTGGATTCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 274
DB 39019 AATTTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 274
QY 275 GTTTTGTGTTCTGTTCTGTTATGTTATGAGTAAATCAAAAAGGATTTGATGAGT 334
DB 39079 ATGTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 334
QY 335 GTAAACATATTCGTTATGACCCCAAAAAAATTAATTAATTAATTAATTAATTAATTA 394
DB 39139 ATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 394
QY 395 CCGATATATGTTTGTGTTCTGTTATGAGTAAATCAAAAAGGATTTGATGAGT 454
DB 39199 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 454

QY 455 TTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 514
DB 39259 AATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 514
QY 515 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 573
DB 39316 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 573
QY 574 TTGCTATATCCAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 633
DB 39376 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 633
QY 634 GAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 693
DB 39436 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 693
QY 694 CCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 753
DB 39496 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 753
QY 754 TTGCTATATCCAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 813
DB 39556 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 813
QY 814 GTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 873
DB 39612 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 873
QY 874 TTGCTATATCCAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 933
DB 39672 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 933
QY 934 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 993
DB 39732 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 993
QY 994 GGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1053
DB 39792 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1053
QY 1054 CATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1113
DB 39852 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1113
QY 1114 GAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1173
DB 39912 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1173
QY 1174 ACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1233
DB 39969 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1233
QY 1234 TTAAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1293
DB 40029 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1293
QY 1294 GAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1353
DB 40088 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1353
QY 1354 TGGACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1413
DB 40148 TTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1413
QY 1414 AAACAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1473
DB 40208 GTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1473
QY 1474 ATCTTGTGTTCTGTTCTGTTATGAGTAAATCAAAAAGGATTTGATGAGT 1533
DB 40268 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1533
QY 1534 TTGCTATATCCAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1593

Db	40328	TATATTTAAATTAATTAATTTTAA-TTATTTTAAATATTTAAATTTAAATTTTATA	40386
Oy	1594	CTTGTAACATTCACATATTATTATATGAGTACAGTATGGATTAT--TCGCAATACATTC	1651
Db	40387	TTTTTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40446
Oy	1652	TTTGGATGTTTAAACCTTAATCTGTCTTCCTACGAGTAATTAATTAATCAATCAGGTA	1711
Db	40447	TTTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40506
Oy	1712	AAAAAGTTTTGCTTATATTTGCGCATGACGACGATTAACCTATATGCTTAATTTT	1771
Db	40507	ATATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40566
Oy	1772	TGAAAATGTAACCTTTTACT---CATAGATTAATTAACCGTAGTGTGTTGGCATTA	1827
Db	40567	ATTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40626
Oy	1828	TGACAGCCTTCACACTGGA-----TAGCAATTTTTTTCGCAATATTTAAATTAAG	1880
Db	40627	TTAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40686
Oy	1881	AATTCAAATGCTACATTCACATAGACAAACAGCTGATATTCATTTTAAATTAAGACA	1940
Db	40687	ATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40746
Oy	1941	AATTTTGGAAAAAGTTTATATTTCTAACAAATTTATTAATTAATTAATTAATTAATTA	2000
Db	40747	ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	40806
Oy	2001	TTTCCATGTCCTTAAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTA	2060
Db	40807	TATATTTAAATTTATTAATTAATTAATTAATTAATTTATTTATTAATTAATTAATTA	40864
Oy	2061	AATAGTGGTGCAATTCAAATATCTCCCAATTAATTTTGGAAATCTACAAATTAATTA	2120
Db	40865	TATATTTAAATTTAAATTTAAATTTAAATTAATTAATTAATTAATTAATTAATTA	40924
Oy	2121	TTTACTGCATTAACATGATAGTAAGGTCGCAAAAAAATTTTGTATACGAAGACTCCA	2180
Db	40925	TTTATTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTTATTTATTA	40984
Oy	2181	ATTTTATTTTATTAATGACAAAGAAATTAACAGATAGAAAACTATTTTGTGGAATGGA	2240
Db	40985	ATATTTTATTAATGTAATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTA	41044
Oy	2241	AGTACTAATTAATCATTAAGCAAAATTTTAAAAATTAATTAATTAATTAATTAATTA	2300
Db	41045	TTTATTTATTTATTAATTAATTAATTTTAAATTAATTTTAAATTAATTAATTTATTA	41104
Oy	2301	ATGTTATCTAGTAGGTGTAATTAATTAATGATGCGATTCGAAATTTGGCAACAATAG	2360
Db	41105	ATTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	41163
Oy	2361	AAAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTGACTA	2407
Db	41164	TATATTTAATTAATTTATTAATTAATTAATTTATTAATTAATTAATTAATTAATTA	41210

RESULT 11	AC120883/c	LOCUS	AC120883	133877 bp	DNA	linear	HTG 06-AUG-2002
DEFINITION			Homo sapiens chromosome 18 clone CTD-2146H24 map 18, WORKING DRAFT SEQUENCE, 4 ordered pieces.				
ACCESSION			AC120883				
VERSION			AC120883.2	GI:22123189			
KEYWORDS			HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE			human.				
ORGANISM			Homo sapiens				
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
			1 (bases 1 to 133877)				

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCES
AUTHORS

Birren,B., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 18, clone CTD-2146H24
 unpublished
 2 (bases 1 to 133877)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
 Boudgater,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A.,
 Cook,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fairo,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,U., Gardyna,S.,
 Ginde,S., Gordon,L., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,S., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazeres,R.,
 Landers,T., Lenockzy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarty,M., McEwan,P., McKernan,K., Meldrum,T., Menues,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunhthang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Sudrmanian,A., Talamas,J., Teeffaye,S., Theodore,J.,
 Topham,K., Travers,M., Trevis,N., Triggillo,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zemek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome
 Research, 330 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 133877)
 Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boudgater,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collimore,A.,
 Cook,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fairo,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
 McCarty,M., Meldrum,T., Menues,L., Mihova,T., Mlenga,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunhthang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Roy,A., Schauer,S., Schnupack,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Teeffaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zemek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (06-MGC-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 6, 2002 this sequence version replaced gi:70531967.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L26769
 Center clone name: 2146_H_24
 Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap, version 0.960731
 Consensus quality: 132654 bases at least Q40
 Consensus quality: 132863 bases at least Q30
 Consensus quality: 133065 bases at least Q20
 Insert size: 13300; agarose-gel
 Insert size: 133577; sum-of-configs
 Quality coverage: 17.1 in Q20bases; agarose-gel

/db_xref="GI:4493932"
/translation="MGLKRVKVFYLVSPSLFAVFLKRRHNNENNYEPLINNDIE
KIKIRHNKCSYIPLEFLNIYDSIYKNNKILRMFYEFRRRRODREXYITNNVRK
KREAIKYNFIDEONLFNKFYIEVLEYSKIGLISPHLSITLKISSECVNITP
SLYNNKLDNHNHNLNEKKIKYFKOINNEHQOAPTNHNNNNKPLDINHSC
KNNYSSTYNNNEKENINHYDKNINNEFTESIYKDNCOHITLMMYILNOTY
DNICICLANTINITYINFMNINILKYICNNMELILANNHIEKKIKINOKNNTS
LFKYISFEFFKKEENHLYDFEODMNNHLKREKDFYNSNEHTNNIKYISNDY
FYDHINSSNRCSEFNKLOOTDNTKHIIGKEKYPNNKSDHEKNNNNKINNIK
DOKDKILKIFYLGNKLDIOILNELVMMNLPFCSCIKLISIKNHILKKE
FDSNDTYNNYDEKTKTNTNNNNNDNDNDCSNNDHCNNNDICSNNDICSN
NNNNICNN
VELAKITNNIFDLKGGKKEHONEDNTINLYIKKFPWIFLYKILNKDSTIEH
NNNIVNGDIKNNNTIKKRAHYIYNFDOFLINNYHIIHKKHKKHIIHKKCKDE
IDIVKFDLYCCMINININISYIHKVDHNCVYRIFKAYKILLERYNLENENY
YKNIKIKYIYLNIN
LYTORIONCODISYIYKAYKFNKNPPLNTLYELHYISEKKRRKKEFSISSP
YSYDPMNSFFSFFSFFSLTFLILFHPDMVASYIEFKLITYSGLPTYYSYNNI
MVCVGPKT
join(7669..7717,7800..7829,7912..7940,8064..8106,
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,
9369..9505,9613..9838)
/gene="PFC0581w"
join(7669..7717,7800..7829,7912..7940,8064..8106,
8302..8368,8568..8641,8713..8812,8924..9006,9122..9258,
9369..9505,9613..9838)
/note="PFC0581w", hypothetical protein, len: 324 aa,
revised: new gene prediction, splicing prediction very
tentative"
/codon_start=1
/product="hypothetical protein, PFC0581w"
/protein_id="Cab90285.1"
/db_xref="GI:7711066"
/translation="MKKKKHYSIKFVNVNDKRYVCSNDYIRIINMIGLHIFRL
OKIFHYFANEOLASSPESHKNRYATKEIDIGIKCMNTIDYIISBKIDNDFS
NDIKTYTAKNKMDOLLTYSNKKIDTYNASFOAOSSEYIFELIKSHRMSGGA
LKYDEIYIKNNFSPALSNDNSVTKYIVDLNLNIISEETKYNFASVGGKVA
TLKREKKILNNRLLSREKYPNNQVMDKKEKRIHITVTIIMLFFLISHHYHDSV
QNFLKEKNNKSDLODDIDEDERKPFDEILREAKKSEYDKDBEL"
7790..7799
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 2 of
PFC0581w"
7830..7835
/gene="PFC0581w"
/note="potential splice donor sequence for exon 2 of
PFC0581w"
7901..7911
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 3 of
PFC0581w"
7941..7947
/gene="PFC0581w"
/note="potential splice donor sequence for exon 3 of
PFC0581w"
8055..8063
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 4 of
PFC0581w"
8107..8112
/gene="PFC0581w"
/note="potential splice donor sequence for exon 4 of
PFC0581w"
8293..8301
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 5 of
PFC0581w"
8369..8374
/gene="PFC0581w"

misc_feature
/note="potential splice donor sequence for exon 5 of
PFC0581w"
8559..8567
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 6 of
PFC0581w"
8642..8647
/gene="PFC0581w"
/note="potential splice donor sequence for exon 6 of
PFC0581w"
8704..8712
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 7 of
PFC0581w"
8813..8818
/gene="PFC0581w"
/note="potential splice donor sequence for exon 7 of
PFC0581w"
8913..8923
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 8 of
PFC0581w"
9007..9012
/gene="PFC0581w"
/note="potential splice donor sequence for exon 8 of
PFC0581w"
9112..9121
/gene="PFC0581w"
/note="potential splice acceptor sequence for exon 9 of
PFC0581w"
9259..9264
/gene="PFC0581w"
/note="potential splice donor sequence for exon 9 of
PFC0581w"
9361..9368
misc_feature
Query Match 3.9%; Score 125.4; DB 3; Length 86827;
Best Local Similarity 44.0%; Pred. No. 2,4e-09;
Matches 1023; Conservative 0; Mismatches 1266; Indels 38; Gaps 10;
QY 108 ATATTAATAATTAATAATTCATTAAATATTTACAGTAATTAATTAATTCATTGCT 167
DB 41351 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 41292
QY 168 ATGTATAACAATAATCATCTATCTGTATGTAATGAAATAATGAGTTGCAATTTATA 227
DB 41291 AATAATAAATACACATCAATTAATAATTAATTAATTAATTAATTAATTAATTAATTA 41232
QY 228 ATATAAAGAAATAATGATTCATTTGTTGGATTACAGTTAAGTTTGTGTTTC 287
DB 41231 ATTTTATATAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 41172
QY 288 TTTTGTATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 347
DB 41171 TAAATATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 41121
QY 348 GTTATGACCCCAAAAAAAAAAAAAAAAAAACAACAACAACCCCCCGGATTAAGTTT 407
DB 41120 AATAATATATATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 41061
QY 408 TTGGTTCTGATAGTTATTTGATCATTAATTAATGATGATGATGATGATGATGATGATG 467
DB 41060 TCATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 41001
QY 468 AGATTTCTTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 524
DB 41000 TAACTATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 40941
QY 525 ACGAGTGTGAATATCCGTTTATCGATCAATCAATGATTAAGATTTGCTGCTATATC 584
DB 40940 TATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 40881
QY 585 CAGCAATTAATTAACAAAGATTTAGAGAAAAACCGAAAAAAGCAAGAGAGTAG 644

[illegible]

Db	39811	ATAAAATATATAAATTAATTTATTT-----TATATATATATATATATTTATTTATTAAC	39760
Oy	1724	TCTTATTTTCGGCAGCAGGAAGATAAACCTATAGACTTTATTTTGGAAAAAGTAC	1783
Db	39739	GCTTATTTTATTCATTTATATATATATATATATATTAATTTGAATTTATTTATTTT	39700
Oy	1784	CGTTTACTCATAGATTAATTAACGGTATGTTTGTGGCATAATGACAGCCTACAAAC	1843
Oy	1844	TGTGATGACGATTTTTCTGCAATATTAATATGTGAATTCGAATGCTACTCTCAATAGA	1903
Db	39659	TATATAAATTTATATATATATAAAAACATTTATTTAT--TTAAATATATTAATTTATATATA	39641
Db	39640	TATATATATATATATTTATTTGCAATATTAATTTATTTATATATATTTATTTATATAT	39581
Oy	1904	AGAAACACCTGAGATTTATTCATTTTATTTTAAAGACAAAATTTTGGAAAAATGTATTAAT	1963
Db	39580	ATATATTAATTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTAATATATANA	39521
Oy	1964	TCTAACCAATATTTTAAAAATGATGCGCTATATATGATATTTCTATGCTCTTAAAAATTT	2023
Db	39520	TTAAATTTATTTTATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTATTTATTT	39461
Oy	2024	TTTTTATATTTAGTTATTAATAACATTTATTAAGCAATATATAGTGGATGATCAATATAC	2083
Db	39460	TATTTTATTTTATTAATAATATATATATATATTTATTTTATTTTATTTAATAATAATATATA	39401
Oy	2084	TCCATTAATATTTTGTGAAATCTACAAATTTATTAATTTTACTGCAATACAAATGCATAGA	2143
Db	39400	TTAATTTATTTTATTTTAAATTAATAATATATATTAATTTATTTATTTTATTTTAAATANA	39341
Oy	2144	AAGTCCAAAAAAATTTTGTTAACAGAACTCCAAATTTTTTTTTTTTATGAAACAG	2203
Db	39340	TATATATTTATTTATTTATTTTATTTTATTAATAATATATATTTATTTATTTTATTTTATTT	39281
Oy	2204	AAATA--ACAGATATGAAAACTATTTTGTGTGATGGAAGTACTAATATATCATTAAGC	2260
Db	39280	AAATATATATATTAATTTATTTTATTTTATTTTAAATAATATATATATTAATTTATTTATTTT	39221
Oy	2261	AAATTTTAAAAATTAATATAGCCATATACGCCCTCAAAAGTATGTATCTAGTAGGTGTA	2320
Db	39220	AATATATATATATTTTATTTTATTTTAAATTAATTTAAAAATTAATTTATTTATTTGTATAT	39161
Oy	2321	TTAA--TAATGATGTCGTCGATTCAGAAATTTGGCAACAATGAAAAACGAATTAATAAAT	2378
Db	39160	TTTATTTTATTAATATATATATATATATTTATTTATTTATTTATATATATTTATATTTAAT	39101
Oy	2379	TACTTTAAATTAATTAATTAATTTTGAGTAATATGTTTCTGCAGCAT	2425
Db	39100	TTAATATATATATATATATACATATTAATTAATTTTATTTTATTTATTTT	39054
RESULT 13			
AC105425/c			
LOCUS	AC105425		
DEFINITION	Homo sapiens chromosome 7 clone RP11-776N17, WORKING DRAFT		
ACCESSION	AC105425		
VERSION	AC105425.2		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	Waterston, R.H.		
REFERENCE	1 (bases 1 to 205130)		
TITLE	The sequence of Homo sapiens clone		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 205130)		
AUTHORS	Waterston, R.H.		
JOURNAL	Submitted (04-JAN-2002) Genome Sequencing Center, Washington		
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
JOURNAL	MO 63108, USA		

QY 903 ACTATATATATCGGCAATATATGATATATTCACAAAACATGCTGGAC 962
 Db 78934 ATCAATAC 78975
 QY 963 CATTAC 1022
 Db 78874 TAT 78815
 QY 1033 ATTTTACAGCCTCAGTATGATATATATATATATATATATATATATAT 1082
 Db 78814 AATGAT 78755
 QY 1083 TTTAT 1142
 Db 78754 TAT 78695
 QY 1143 GGTGCTAT 1202
 Db 78694 TTTAT 78636
 QY 1203 AT 1262
 Db 78635 TAT 78576
 QY 1263 GAT 1322
 Db 78575 AT 78516
 QY 1323 TAAATGAAATTCACAGCGGTATATATATATATATATATATATATATAT 1382
 Db 78515 TAT 78466
 QY 1383 ATCATCTGTGCTGTCACATATATATATATATATATATATATATATAT 1442
 Db 78465 TTTAT 78406
 QY 1443 TGGATTCATTCATGATATATATATATATATATATATATATATATATATAT 1502
 Db 78405 TAT 78346
 QY 1503 TACAAGCTGTATACCATGCAATATATATATATATATATATATATATATAT 1562
 Db 78345 TACAT 78287
 QY 1563 TTATTTGATAGTGGCAATTAATATATATATATATATATATATATATATAT 1622
 Db 78286 AT 78227
 QY 1623 ACTAT 1682
 Db 78226 CAT 78167
 QY 1683 CTACGCTAT 1742
 Db 78166 TAT 78107
 QY 1743 GAAGAT 1802
 Db 78106 AT 78048
 QY 1803 TTACGCTAT 1862
 Db 78047 AT 77988
 QY 1863 TGCAT 1922
 Db 77987 TAT 77928
 QY 1923 ATTTTAT 1982
 Db 77927 AT 77868
 QY 1983 TAT 2042

Db 77867 TAT 77808
 QY 2043 AAT 2102
 Db 77807 TAT 77749
 QY 2103 ATCTACAAAT 2162
 Db 77748 TTTAT 77689
 QY 2163 GTTACAGAACTTCCAAATTTTATATATATATATATATATATATATATAT 2222
 Db 77688 TAT 77629
 QY 2223 ATTTTGTGGAATGAGATATATATATATATATATATATATATATATATATAT 2282
 Db 77628 TTTAT 77569
 QY 2283 CCTATACGCGCTCAAGATATATATATATATATATATATATATATATATATAT 2342
 Db 77568 AT 77509
 QY 2343 AGAATGGACACAT 2402
 Db 77508 ATCTAAAT 77449
 QY 2403 GAGTAAATGTGTTCTGACTAT 2425
 Db 77448 TAT 77426

RESULT 14
 DMU11584
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

DMU11584 4601 bp DNA linear INV 23-JUL-1994
 Drosophila melanogaster Oregon-R mitochondrial A+T region.
 U11584
 U11584.1 GI:508826
 mitochondrial DNA; A+T region; tandem repeats.
 fruit fly.
 Mitochondrion Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 4601)
 Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
 Sequence, organization, and evolution of the A+T region of
 Drosophila melanogaster mitochondrial DNA
 Mol. Biol. Evol. 11 (3), 523-538 (1994)

JOURNAL
 MEDLINE
 PUBMED
 8015445
 2 (bases 1 to 4601)
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Direct Submission
 Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D. Dept. of
 Biochemistry, Michigan State University, East Lansing, MI,
 48824-1318, USA

FEATURES
 source

1. 4601
 /organism="Drosophila melanogaster"
 /organelle="mitochondrion"
 /strain="Oregon R"
 /db_xref="taxon:7227"
 /dev_stage="embryo"
 650..1022
 /note="repeat I-A"
 /rpt_type=tandem
 1023..1360
 /note="repeat I-B1"
 /rpt_type=tandem
 1361..1705
 /note="repeat I-C/A"
 /rpt_type=tandem

repeat_unit
 repeat_unit
 repeat_unit

Db 3941 ATTATGAAATATTTATATATATATATAT-ATATATAGAAAAATTAATTTAATTA 3999
 QY 1787 TTTACTCATAGATTAATTAACGGATGTTTGGTCCATTAAGACGGCTACAACTGT 1846
 Db 4000 TTTAATATATAATTTTAAAAATTTCTTAATATGATATATTTTAAAAAATATTTATAT 4059
 QY 1847 GATAGTCAATTTTTCGCAATATTAATTAATGCAATTCATGCTATCATCAAGAA 1906
 Db 4060 AATAAATCAATTTTAAAAAATTAACAAATTTTATATTAATTAATTTTAAAG 4119
 QY 1907 AACACCTGAGTATTCATTTTATTAAGCAAAATTTTGAATAATGTTAATTTCT 1966
 Db 4120 AATATATATTTATTTATTTTATTTTATTTTAAAAAATTTTAAAAAATTTTATTT 4179
 QY 1967 AACATATTTATTAATTAATGATGCTATTAATGATATTTCTTAATTAATTTT 2026
 Db 4180 TTTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4239
 QY 2027 TTTATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 2086
 Db 4240 TTAATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAG 4299
 QY 2087 ATTAATATTTTGAATTCATTAATTTATTAATTTATGTCATTAACATGATGAAG 2146
 Db 4300 ATATATATTTATTAATTAATTAATTAATTAATTAATTTATTAATTTATTAATTAAT 4359
 QY 2147 TTCCAAAAAATTTTGTAAACAGAACTCCAAATTTTAAAAATTTTAAAGAAAGAA 2206
 Db 4360 TAAATATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4419
 QY 2207 TAACGATGAGAACTATTTTGTGTGAATGGAATGATTAATTAATTAATTAATTAAT 2266
 Db 4420 TATATATATTAATTAATTAATTAATTAATTAATTAATTAATTTTCAATTAATTAAT 4479
 QY 2267 TAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2326
 Db 4480 AAAAAATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4539
 QY 2327 ATGATGTCGATTCAGATTTGGACAAACATGAAGAAATTAATTAATTAATTTA 2386
 Db 4540 TTTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4599

RESULT 15

DM037541

LOCUS 19517 bp DNA circular INV 02-MAR-2001
 DEFINITION Drosophila melanogaster complete mitochondrial genome.
 ACCESSION U37541
 VERSION U37541.1 GI:1166529

SOURCE

ORGANISM Drosophila melanogaster.
 Mitochondrion Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 12511 to 12682)
 Clary,D.O., Goddard,J.M., Martin,S.C., Faumon,C.M. and
 Wolstenholme,D.R.

REFERENCE

JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)
 MEDLINE 83090428
 PUBMED 6294611

REFERENCE

JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)
 MEDLINE 8320794
 PUBMED 6304652

AUTHORS 3 (bases 404 to 5272)
 de Bruijn,M.H.
 TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and

JOURNAL genetic code
 MEDLINE Nature 304 (5923), 234-241 (1983)
 PUBMED 83245048
 4 (bases 804 to 1778)

REFERENCE

AUTHORS Satta,Y., Ishiwa,H. and Chigusa,S.I.
 TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in
 Drosophila melanogaster and its sibling species
 MOL. BIOL. EVOL. 4 (6), 638-650 (1987)
 JOURNAL 81174373
 MEDLINE 88212147
 PUBMED 2832697

REFERENCE

AUTHORS Garesse,R.
 TITLE Drosophila melanogaster mitochondrial DNA: gene organization and
 evolutionary considerations
 GENETICS 118 (4), 649-663 (1988)
 JOURNAL 3130291
 MEDLINE 88212147
 PUBMED 3130291

REFERENCE

AUTHORS Satta,Y. and Takahata,N.
 TITLE Evolution of Drosophila mitochondrial DNA and the history of the
 melanogaster subgroup
 Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
 JOURNAL 9108857
 MEDLINE 2124697
 PUBMED 2124697

REFERENCE

AUTHORS Ballard,J.W., Olsen,G.J., Falch,D.P., Odgers,W.A., Rowell,D.M. and
 Atkinson,P.W.
 TITLE Evidence from 12S ribosomal RNA sequences that onychophorans are
 modified arthropods
 Science 258 (5086), 1345-1348 (1992)
 JOURNAL 93088057
 MEDLINE 1455227
 PUBMED 1455227

REFERENCE

AUTHORS Lewis,D.L., Farr,C.L., Farguhar,A.L. and Kaguni,L.S.
 TITLE Sequence, organization, and evolution of the A+T region of
 Drosophila melanogaster mitochondrial DNA
 MOL. BIOL. EVOL. 11 (3), 523-538 (1994)
 JOURNAL 94285822
 MEDLINE 8015445
 PUBMED 8015445

REFERENCE

AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
 TITLE Drosophila melanogaster mitochondrial DNA: completion of the
 nucleotide sequence and evolutionary comparisons
 Insect MOL. BIOL. 4 (4), 263-278 (1995)
 JOURNAL 96423163
 MEDLINE 8825764
 PUBMED 8825764

REFERENCE

AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
 TITLE Direct Submission
 Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
 Michigan State University, East Lansing, MI 48824-1319, USA
 JOURNAL location/Qualifiers

FEATURES

1..19517
 /organism="Drosophila melanogaster"
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /note="derived from new and previously submitted
 sequences; sequence is a composite containing sequences
 obtained from different Drosophila melanogaster strains"
 1..65
 /product="trna-ile"
 /complement(97..165)
 /product="trna-gln"
 171..239
 /product="trna-phe"
 240..1265
 /codon_start=1
 /transl_table=5
 /product="NADH dehydrogenase subunit 2"
 /protein_id="AAC47810.1"
 /db_xref="GI:1166530"
 /translation="MFNNSKILFTTMIIGTLITVTSNGLGAMGLEINLLSFIPL"


```

Db 1777 ATATCTAATATTTAAATTTAAATTTAAATTTAAATTTAGATATATTTATA 17836
QY 724 AGCTTCCTCACAATCATATGACCGCTATTTGGATTTTCGATATTCGTTAAATTT 783
Db 17837 AAATTAATATCTCATATTTATTTATTTATTTAAATTTATATTAATTAATTAATTT 17896
QY 784 GCGATACGATGTGAAA--ATATTTTATTTGTAGTCGATCTCAATTTATGT-TCGA 840
Db 17897 TAATTAATTTATTTATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTA 17956
QY 841 GGTATTCGATATCTTCGTTTAAAGCATATTTGTCTTTCTTTCTTTCTTTCTTC 900
Db 17957 TATATATGAAAATTTAATTTATTTAAATTTAATTTAATTTAATTTTAAATTTCTTA 18016
QY 901 TAACATATATTTATCGCGGATATAT---GATTAACATGATATATACCAAAACATTTGT 955
Db 18017 AATGATATTTATTTTAAATTTATTTATTTATTTAAATTTCAATTTTAAATTTAAAC 18076
QY 956 CTGGACCATTTTGAATAAATTTCTTCAAAACATTTACGGGACACTGACCTGACCCCTTA 1015
Db 18077 AAAAAATTTTAAATTTAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTTTCAATTT 18136
QY 1016 AAATACGATTTTACAGCTCAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 1075
Db 18137 TTAATAAATTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 18196
QY 1076 CA----ACGATTTTAAACAAAGCTTCAAACTGAATATAGCTTGAATCTTTAGAAAT 1130
Db 18197 AATTAATGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18256
QY 1131 TTTGAATTTACCGTGTGTATGTATTAATATAGATTTAGTGTAACAAATATCTTAATCA 1190
Db 18257 AAATTAATAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTCTCAT 18316
QY 1191 TTAGTGTCACAAATATATATTTCTTTACAGAAAAACAAATTTAGAGAGATTTACAT 1250
Db 18317 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18376
QY 1251 ATTCATATATGAGGATATGCTATACCTTTACAGCTAGCTATAGTAAAGAAATAGTT 1310
Db 18377 AATTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18436
QY 1311 ATGATGTCGATTAATGAATTTACAGCGCTGATTAATTT--ATGGACCGCATGTT 1367
Db 18437 AAATTTATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAT 18496
QY 1368 ACATCAGCTGCAAAATATCTCTGTTGTCACAAATTTAAACAAAAACAAGAAAAA 1427
Db 18497 AAAAAATTTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18556
QY 1428 GAAAAAGATTTTCTGCAATTCATCAATGATCTAAATGCAATGCAATCTTTGGGTTAC 1487
Db 18557 ATTAATTTTATATGAAATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18616
QY 1488 AGTTGCAAGTCCTCTACAAGGCTGTAACGCTGCAACTATTAATTTGCTTCT-TTAA 1546
Db 18617 AAAAAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 18676
QY 1547 TGCATCTTTAATATTTATTTAGTTGGAATTTAATTAAGAGCACTTTGTAACATTAC 1606
Db 18677 TATATATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18736
QY 1607 AATATTTATTTAGATAGTATGATTTATTCGAAATATACATCTTTGGATGTTTAAAC 1666
Db 18737 AATTTAAATTTAGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 18796
QY 1667 TTAATCTGTTCTCTCCACGCTATTAATTTAATCATCGAGTAAAAAGTTTGTCT 1726
Db 18797 TTTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 18856
QY 1727 TATTTTCGAGATGCAAGGATTAACCTAATGACTTTAATTTTGAATGTAACCT 1786
Db 18857 ATTAATGAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 18915

```

```

QY 1787 TTTACTATAGATTAATTAATTTACCGTATGTTTGTGCCAATATGACAGCCTCAACAATGT 1846
Db 18916 TTTAATATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 18975
QY 1847 GATAGCAATTTTCTCGAATATTAATTTAGAAATTTCAAGCTCTATCATTAAGAA 1906
Db 18976 AATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 19035
QY 1907 AACAGCTGAGTATTAATTAATTTAATTAAGCAAAATTTTGAATAATGTTATATTTCT 1966
Db 19036 AATTAATTAATTTTATTTATTTTCTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 19095
QY 1967 AACATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 2026
Db 19096 TTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 19155
QY 2027 TTTATATTTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2086
Db 19156 TTTAATTAATTTATTAATTAATTAATTAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 19215
QY 2087 ATTAATATTTTGAATTTCAAAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAAT 2146
Db 19216 ATTAATATTTAATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAAT 19275
QY 2147 TTTCAAAAAATTTTGTAAACAGAAATTTCCAAATTTTATTTTATTAAGAACAGAA 2206
Db 19276 TATATTAATTAATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAAT 19335
QY 2207 TAACAGATGAAGAACTATTTTGTGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 2266
Db 19336 TATTAATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAAT 19395
QY 2267 TAAAAATTTATTAATTTAAAGCTTACGCGCTCAAGATGATCTAGTATGATGATTAATTAATTA 2326
Db 19396 AAAAAATTTCTTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAAT 19455
QY 2327 ATGATGTTGCAATTTGCAATTTGGCAACAATGAAACGAAATTAATTAATTAATTTA 2386
Db 19456 TTTTAAATTTAAATTTAGTAAATTAATTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 19515

```

Search completed: March 30, 2003, 08:19:09
 Job time : 34395.6 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:05:35 ; Search time 639.576 Seconds

(without alignments)

11274.505 Million cell updates/sec

Title: US-09-502-426a-1_COPY_1_3202

Perfect score: 3202

Sequence: 1 atgtgggtattatattgtg.....gagagagaagaactagctcc 3202

Scoring table: IDENTITY_NNC

Gapop 10.0 , Gapept 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3202	100.0	6888	21	AAA59599
2	84.6	2.6	883	22	AA15210
3	84.6	2.6	9810	24	ABL32427
4	83.8	2.6	8170	24	ABK28257
5	82.8	2.6	6131	24	ABL32890
6	82.6	2.6	9810	24	ABL32426
7	82.6	2.6	15732	22	AA545388
8	82.6	2.6	15732	24	ABK28233
9	81.6	2.5	5689	22	AA545384

C	10	81.6	2.5	5689	22	AA546426	Tumour suppressor
C	11	81.6	2.5	5689	24	ABK28226	DNA transcription
C	12	80.6	2.5	6294	24	ABL33054	Human immune syste
C	13	80.4	2.5	6131	24	ABL32891	Human immune syste
C	14	80	2.5	640681	24	ABA92787	Buchera sp. genom
C	15	79.6	2.5	47108	24	ABK31511	Signal transductio
C	16	79.4	2.5	883	22	AA15210	Human breast cance
C	17	79.4	2.5	8310	20	AA229911	CDNA encoding a SC
C	18	79.2	2.5	18283	24	ABL70502	Human gene regulat
C	19	78.2	2.4	50000	24	AA561363	Human immune syste
C	20	78.2	2.4	17137	24	ABL55643	Human immune syste
C	21	77.8	2.4	17137	24	ABL32191	Human immune syste
C	22	77.6	2.4	8810	24	ABL32427	Human immune syste
C	23	77.4	2.4	19124	18	AA172882	Plasmodium var-7 p
C	24	77.4	2.4	19124	21	AA298287	Plasmodium var-7 p
C	25	77.2	2.4	15373	24	ABL32467	Human immune syste
C	26	77	2.4	11691	24	ABL34241	Human immune syste
C	27	76.8	2.4	50000	24	ABL55643	Human immune syste
C	28	76.4	2.4	10234	24	ABL33075	Human immune syste
C	29	76.2	2.4	15348	24	ABL34155	Human immune syste
C	30	76.2	2.4	50000	24	ABL56202	Human immune syste
C	31	76	2.4	6050	24	ABL34011	Human immune syste
C	32	75.6	2.3	5979	22	AA545313	Human immune syste
C	33	75.2	2.3	5979	22	ABK28152	Human immune syste
C	34	75.2	2.3	38342	22	AA546746	Human immune syste
C	35	75	2.3	38342	22	ABK31507	Human immune syste
C	36	75	2.3	6963	24	ABL32979	Human immune syste
C	37	74.8	2.3	10710	24	ABL32893	Human immune syste
C	38	74.8	2.3	11422	24	ABK39936	Human immune syste
C	39	74.8	2.3	11422	24	ABK32218	Human immune syste
C	40	74.8	2.3	10250	24	ABN80083	Human immune syste
C	41	74.4	2.3	700	22	AAH93026	Human chemically m
C	42	74.2	2.3	6641	24	ABL54336	Human inflammatory
C	43	74.2	2.3	6641	24	ABL32315	Chemically treated
C	44	74.2	2.3	2104	13	AA025273	Human immune syste
C	45	74	2.3				Sequence encoding

ALIGNMENTS

RESULT 1	AAA59599	standard; DNA; 6888 BP.
ID	AAA59599	
XX	AAA59599:	
AC	14-NOV-2000	(first entry)
XX		
DE	DNA encoding a cytochrome P450 enzyme designated DMF4.	
XX		
KW	DMF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;	
KW	plant phenotype; cell elongation; ss.	
XX		
OS	Arabidopsis sp.	
XX		
PH	Key	Location/Qualifiers
FT	Promoter	1..3203
FT	TATA_signal	3060..3125
FT		/*tag- a
FT		3203..3423
FT		/*tag- b
FT	CDS	3203..6110
FT		/*tag- c
FT		/*product- "DMF4"
FT		/*note- "contains introns"
FT	exon	3203..3423
FT		/*tag- d
FT	intron	3424..3503
FT		/*tag- e
FT	exon	3504..3828
FT		/*tag- f
FT	intron	3829..3913
FT		/*tag- g

FT exon 3914..4066
 FT /tag- h
 FT Intron 4067..4164
 FT /tag- i
 FT exon 4165..4479
 FT /tag- j
 FT Intron 4480..4631
 FT /tag- k
 FT exon 4632..4724
 FT /tag- l
 FT Intron 4725..4815
 FT /tag- m
 FT exon 4816..4894
 FT /tag- n
 FT Intron 4895..5000
 FT /tag- o
 FT exon 5001..5110
 FT /tag- p
 FT Intron 5111..5864
 FT /tag- q
 FT exon 5865..6110
 FT /tag- r
 FT 3'UTR 6011..6468
 FT /tag- s
 PN WO20047715-A2.
 XX 17-AUG-2000.
 PD 11-FEB-2000; 2000WO-US03820.
 PE 11-FEB-1999; 99US-0119657.
 PR 11-FEB-1999; 99US-0119658.
 XX
 XX (ARIZ-) ARIZONA BOARD OF REGENTS.
 PA
 PI Aspiroz R, Choe S, Feldmann KA;
 XX WPI: 2000-549142/50.
 DR P-PSDB: AAB07921.
 XX
 PT New isolated dwf4 polynucleotide useful for altering the phenotype of
 XX plants, for diagnostic assays and in the production of antibodies -
 PS
 PS Claim 3; Fig 10A-G; 113pp; English.
 CC The present sequence encodes a DMF4 polypeptide. The polypeptide is a
 CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
 CC brassinosteroids. Specifically, it mediates multiple
 CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DMF4
 CC polynucleotide is used for altering the phenotype of a plant. DMF4
 CC plants display a dramatic reduction in the length of different organs,
 CC and this size reduction is attributable to a defect in cell elongation.
 CC The DMF4 polynucleotides and polypeptides can be used in diagnostic
 CC assays and to generate antibodies, which can be used to produce
 CC immunogenic compositions.
 XX
 XX Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;
 S0
 Query Match 100.0%; Score 3202; DB 21; Length 6888;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 AAAATTCATTAAAAATTTTACAGTAATTAATTAATCTTACATCTGATGTATTAACA 180
 Qy 181 AATATCTATCTTGGATATGAAAAATATGAGTTGAAATTTAATAATAAGGAAA 240
 Db 181 AATATCTATCTTGGATATGAAAAATATGAGTTGAAATTTAATAATAAGGAAA 240
 Qy 241 TAATCGATTCATTGGTTGGATTACACAGTAAGTTTGTGTTTCTTTGTATATGT 300
 Db 241 TAATCGATTCATTGGTTGGATTACACAGTAAGTTTGTGTTTCTTTGTATATGT 300
 Qy 301 ATATGAGTAATCAAAAAAGATTTGATGAAAGTGTAACAATTTCCATTAGCCCCA 360
 Db 301 ATATGAGTAATCAAAAAAGATTTGATGAAAGTGTAACAATTTCCATTAGCCCCA 360
 Qy 361 AAAAAAACAACAACAACAACCCGCCGATAGTTTGTGTTGGATT 420
 Db 361 AAAAAAACAACAACAACAACCCGCCGATAGTTTGTGTTGGATT 420
 Qy 421 AGTTTATTTGATCATTAATTACATGCATCTTCTTGTACTATGAAGATTTCTTAC 480
 Db 421 AGTTTATTTGATCATTAATTACATGCATCTTCTTGTACTATGAAGATTTCTTAC 480
 Qy 481 CAATTAATTTTGCATTTGATATATCTCTGATTTATTAATTAATACAGATGGAATCC 540
 Db 481 CAATTAATTTTGCATTTGATATATCTCTGATTTATTAATTAATACAGATGGAATCC 540
 Qy 541 GTTATCGATCACTCCATCATGATTAATGATCTGTGTGTAATCCAGCAATTAATTA 600
 Db 541 GTTATCGATCACTCCATCATGATTAATGATCTGTGTGTAATCCAGCAATTAATTA 600
 Qy 601 AGATTAATGAGAAAAACCGAAAAATAGAAAAAGGAAAGAGTAGACCATGAGATAG 660
 Db 601 AGATTAATGAGAAAAACCGAAAAATAGAAAAAGGAAAGAGTAGACCATGAGATAG 660
 Qy 661 TGAATTAATTAATCAAGAAATAGAGATGACAAACCAAAAGTTGTGATATGTCCT 720
 Db 661 TGAATTAATTAATCAAGAAATAGAGATGACAAACCAAAAGTTGTGATATGTCCT 720
 Qy 721 GCCAGCTTCTCTCAATCAATATCGAACCCATTTGATTTGATTTGATTTGATTTA 780
 Db 721 GCCAGCTTCTCTCAATCAATATCGAACCCATTTGATTTGATTTGATTTGATTTA 780
 Qy 781 TTTGGATTAACGATTTGAAAAATTTTATTTGTTAGCTGATCAATATATATGTCCA 840
 Db 781 TTTGGATTAACGATTTGAAAAATTTTATTTGTTAGCTGATCAATATATATGTCCA 840
 Qy 841 GGTATTTGATTAATCTTCTGTTAAAGCATTTTGTCTTTTGTGTTGTTCTCT 900
 Db 841 GGTATTTGATTAATCTTCTGTTAAAGCATTTTGTCTTTTGTGTTGTTCTCT 900
 Qy 901 TAATATATATTTATGCGGATATATGATTAATATATATATATATATATATATAT 960
 Db 901 TAATATATATTTATGCGGATATATGATTAATATATATATATATATATATATAT 960
 Qy 961 ACCATTTGATTAATCTTTTCTCAAACTTTACGGGACCTGACCTGACCTTAAATA 1020
 Db 961 ACCATTTGATTAATCTTTTCTCAAACTTTACGGGACCTGACCTGACCTTAAATA 1020
 Qy 1021 CGATTTACAGGCTACATAGTATGATTAATGATTAATGATTAATGATTAATGATTA 1080
 Db 1021 CGATTTACAGGCTACATAGTATGATTAATGATTAATGATTAATGATTAATGATTA 1080
 Qy 1081 TATTTATCAAAAGTTACAACAGTAATATAGTTGAATCTTTAGAAAATTTTGGAAAT 1140
 Db 1081 TATTTATCAAAAGTTACAACAGTAATATAGTTGAATCTTTAGAAAATTTTGGAAAT 1140
 Qy 1141 CCGGTTGTTATGTAATATATAGTTTATGTTGTAACAAATATGTTATCAATATAGTCT 1200
 Db 1141 CCGGTTGTTATGTAATATATAGTTTATGTTGTAACAAATATGTTATCAATATAGTCT 1200
 Qy 1201 ACATATACATATCTTCAAGAAAAACAACCTTAAGGAAGTTAATCAATCAATATAT 1260
 Db 1201 ACATATACATATCTTCAAGAAAAACAACCTTAAGGAAGTTAATCAATCAATATAT 1260

```

QY 1261 GGGTATGCTATACCTTTCACGTAATGCTATACGAGACTAAGAAATAGTATGATGTC 1320
    |||||
Db 1261 GGGTATGCTATACCTTTCACGTAATGCTATACGAGACTAAGAAATAGTATGATGTC 1320
QY 1321 GATTAATGAAATGCACACGGGTGTATATATGAGACCTGATGTAGACAGCA 1380
    |||||
Db 1321 GATTAATGAAATGCACACGGGTGTATATATGAGACCTGATGTAGACAGCA 1380
QY 1381 ATATCATCTCTGGTGTGCACATATAAACAAGAAAAAAGAAAAAGATTTT 1440
    |||||
Db 1381 ATATCATCTCTGGTGTGCACATATAAACAAGAAAAAAGAAAAAGATTTT 1440
QY 1441 CTTGATTCATCAATGATCAATTAATGATGATCTTTGGGTACAGTTGGAAGTCC 1500
    |||||
Db 1441 CTTGATTCATCAATGATCAATTAATGATGATCTTTGGGTACAGTTGGAAGTCC 1500
QY 1501 TCTACAAAGCTGTACACATCTGACATTAATGCTTCTTAATGATCTTTAAT 1560
    |||||
Db 1501 TCTACAAAGCTGTACACATCTGACATTAATGCTTCTTAATGATCTTTAAT 1560
QY 1561 ATTTATGTTAGTGTGAAATTAATAGAGCACTTGAACATTAATTAATATAG 1620
    |||||
Db 1561 ATTTATGTTAGTGTGAAATTAATAGAGCACTTGAACATTAATTAATATAG 1620
QY 1621 ATACATGATGATGATTAATCCAAATACATCTTGAATGTTAACTTAATCTGTTCT 1680
    |||||
Db 1621 ATACATGATGATGATTAATCCAAATACATCTTGAATGTTAACTTAATCTGTTCT 1680
QY 1681 TCTACAGGCTAAATATATATATCATCGAGTAAGAAAAAGTTTGTCTTATTTGGCAGTC 1740
    |||||
Db 1681 TCTACAGGCTAAATATATATATCATCGAGTAAGAAAAAGTTTGTCTTATTTGGCAGTC 1740
QY 1741 ATGAAGATTAACCTATGACTTAAATTTTGAATAATGAACCTTATCTCATAGAT 1800
    |||||
Db 1741 ATGAAGATTAACCTATGACTTAAATTTTGAATAATGAACCTTATCTCATAGAT 1800
QY 1801 AATTAACGATGTTTGTGTCATTAATGACACGCTCTACACTGTGATGTCATTTT 1860
    |||||
Db 1801 AATTAACGATGTTTGTGTCATTAATGACACGCTCTACACTGTGATGTCATTTT 1860
QY 1861 TCTGCAAAATTAATTAATAGCAATCTACTATCAATAGAAAGCAAGCAGATAT 1920
    |||||
Db 1861 TCTGCAAAATTAATTAATAGCAATCTACTATCAATAGAAAGCAAGCAGATAT 1920
QY 1921 ACATTTTAATTAAGCAAAATTTTGAATAATGTAATTTCTACATATTAATTA 1980
    |||||
Db 1921 ACATTTTAATTAAGCAAAATTTTGAATAATGTAATTTCTACATATTAATTA 1980
QY 1981 AATATGATGCTATATATGATGTTCTATGTTCTAAATATTTTATTAATTAAGT 2040
    |||||
Db 1981 AATATGATGCTATATATGATGTTCTATGTTCTAAATATTTTATTAATTAAGT 2040
QY 2041 TAATATCATATATAGCAAAATAGTGTGATTCAAATATCTCCATTAATTTTGG 2100
    |||||
Db 2041 TAATATCATATATAGCAAAATAGTGTGATTCAAATATCTCCATTAATTTTGG 2100
QY 2101 AATATCAAAATATTAATATTAATAGTCAATTAATGCAATGAGAAAGTCCAAAAAAT 2160
    |||||
Db 2101 AATATCAAAATATTAATATTAATAGTCAATTAATGCAATGAGAAAGTCCAAAAAAT 2160
QY 2161 TTGTTAAACAGAACTTCACAAATTTTATTAAGAACAGAAATTAAGATAGAAA 2220
    |||||
Db 2161 TTGTTAAACAGAACTTCACAAATTTTATTAAGAACAGAAATTAAGATAGAAA 2220
QY 2221 CTATTTTGTGTGGAATGAGATGATATATATATTAAGAAATTTTAAATATATA 2280
    |||||
Db 2221 CTATTTTGTGTGGAATGAGATGATATATATATTAAGAAATTTTAAATATATA 2280
QY 2281 AGCCTATACGCGCTCAAAAGATGTTATCTAGTGTGATTAATTAATGATGCTGAT 2340
    |||||
Db 2281 AGCCTATACGCGCTCAAAAGATGTTATCTAGTGTGATTAATTAATGATGCTGAT 2340

```

```

QY 2341 TCAGAAATGGGACACATGAAAAAGCAATTAATTAATTAATTAATTAATTAAT 2400
    |||||
Db 2341 TCAGAAATGGGACACATGAAAAAGCAATTAATTAATTAATTAATTAATTAATTAAT 2400
QY 2401 TTGAGTAAATGTGTTTCTGACTATTTAGGGGCAAAAAAAGCAATGCCAAAGTCTAC 2460
    |||||
Db 2401 TTGAGTAAATGTGTTTCTGACTATTTAGGGGCAAAAAAAGCAATGCCAAAGTCTAC 2460
QY 2461 GGGTTTGACTGTCTGACATTCGATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
    |||||
Db 2461 GGGTTTGACTGTCTGACATTCGATTAATTAATTAATTAATTAATTAATTAATTAAT 2520
QY 2521 TAGGGGTCTCTGACATTTTCACTGTCTACCCCTACTGCTGAGCCACCTTTTCCA 2580
    |||||
Db 2521 TAGGGGTCTCTGACATTTTCACTGTCTACCCCTACTGCTGAGCCACCTTTTCCA 2580
QY 2581 TATCTTAAGGTAATTTTGGAAATCCCAATTAATTAATTAATTAATTAATTAATTAAT 2640
    |||||
Db 2581 TATCTTAAGGTAATTTTGGAAATCCCAATTAATTAATTAATTAATTAATTAATTAAT 2640
QY 2641 TGGGATCTGCTGAGCATTTATCAAAATTAATTAATTAATTAATTAATTAATTAAT 2700
    |||||
Db 2641 TGGGATCTGCTGAGCATTTATCAAAATTAATTAATTAATTAATTAATTAATTAAT 2700
QY 2701 AACTCACAACTGATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
    |||||
Db 2701 AACTCACAACTGATCAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2760
QY 2761 ATCTAATGACTTTTCTTCTACACAGGTGATGAAAGTATTAATTAATTAATTAATTAAT 2820
    |||||
Db 2761 ATCTAATGACTTTTCTTCTACACAGGTGATGAAAGTATTAATTAATTAATTAATTAAT 2820
QY 2821 ACAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
    |||||
Db 2821 ACAATGATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2880
QY 2881 CTATTTTCAGATGCGACCTTCTGCAACTTTTGTGTTTAAATTAATTAATTAATTAAT 2940
    |||||
Db 2881 CTATTTTCAGATGCGACCTTCTGCAACTTTTGTGTTTAAATTAATTAATTAATTAATTAAT 2940
QY 2941 AAGTATTAAGAGCAATTAAGAGGCAAAAGTAATTAATTAATTAATTAATTAATTAAT 3000
    |||||
Db 2941 AAGTATTAAGAGCAATTAAGAGGCAAAAGTAATTAATTAATTAATTAATTAATTAATTAAT 3000
QY 3001 TGAAGCTCATTTGTTAGTATTAAGTATTAAGAGTATTAATTAATTAATTAATTAATTAAT 3060
    |||||
Db 3001 TGAAGCTCATTTGTTAGTATTAAGTATTAAGAGTATTAATTAATTAATTAATTAATTAATTAAT 3060
QY 3061 ATAACAATTAATTTTCTGACTTCTTAAACCCCTCTTACAAACAGAAAGCTCCCTTTT 3120
    |||||
Db 3061 ATAACAATTAATTTTCTGACTTCTTAAACCCCTCTTACAAACAGAAAGCTCCCTTTT 3120
QY 3121 TCAGTAAAGTCCGATTCCTTAATTAAGCAAAAGCATTAAGAAAGAAAGTGA 3180
    |||||
Db 3121 TCAGTAAAGTCCGATTCCTTAATTAAGCAAAAGCATTAAGAAAGAAAGTGA 3180
QY 3181 GAGAGAGAGAAAGTACGCTCC 3202
    |||||
Db 3181 GAGAGAGAGAAAGTACGCTCC 3202

```

RESULT 2
AAL15210/c
ID AAL15210 standard; cDNA; 883 BP.
XX
AC AAL15210;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 7667.
XX
KW Human; breast cancer; cell marker; cytosstatic; ss.
XX

CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX
 SQ Sequence 9810 BP; 2961 A; 79 C; 2101 G; 4669 T; 0 other;

Query Match 2.6%; Score 84.6; DB 24; Length 9810;
 Best Local Similarity 44.0%; Pred No. 0.00015;
 Matches 407; Conservative 0; Mismatches 514; Indels 4; Gaps 1;

OY 1527 TATTAATGCTTCTTATATGACCTTATATTTATTTAGTGAATTAATA 1586
 DB TTTAAAGGAGTACGTGAAATATATGATTAATAAATATATATTTATAT 7905
 OY 1587 GAGCGAAGCTTGACATTTATATTTATGATGATGATTTTCCAAATA 1646
 DB TATTTTAGAGTAAATTTTATTTTATTAAGTTATTAAGTAAATTTGAAT 7965
 OY 1647 CAACTTGGATGCTTAACTATCTGTTCTCCAGTAAATATTAATCATCG 1706
 DB TATTTAGAGAAATAGAGAAATGATGAAAGATTTGTTATTTTATTTGAA 8025
 OY 1707 AGGTAAAAAGTTGCTTATTTTCCGATGATGAAGATTAAGATTTAA 1766
 DB TATTTAGAGTAAATTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 8085
 OY 1767 TTTTGGAAATGTAACCTTTTACTCATGATTAATACCGATGTTTGTCC 1826
 DB TTTTATTTGAAATTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 8085
 OY 8086 TTTTATTTAAAAATTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTG 8145
 OY 1827 ATGACAGCCTTCAACGTGATGATCAATTTTCTGCAATATTAATAGA 1886
 DB TGAATATTTATTAATATATAGATTAAGATTTTGGTTTACGAAAAATG 8205
 OY 1887 ATGCTACTATCAATAGAGAAACAGCTGATTTACATTTTAAAGCA 1946
 DB AATTGAGAGATGATGATTAATAAATATATATTTTATTAATTTTATTA 8265
 OY 1947 TGAATAATGTTATTTTCTACAAATTTTAAATATGATGATTTCC 2006
 DB TATATATGTTATATTTATTAATAAATATATATTTTAAATATTTTATA 8325
 OY 2007 ATGCTTAAAAATTTTATTTTATTTAGTATTAATACATTAGAACCA 2066
 DB TATATGTTATATATATATGTTATATATATATATATATATATATATAT 8385
 OY 2067 TGGTGAATCAATATCTCCATTAATTTTGAATCTCAAAATTAATTA 2126
 DB TGTATATATATATGTTATATATATATATATATATATATATATATAT 8445
 OY 2127 CATATACATGCTAGAAAGTTCCAAAAATTTTGTACGAAACTCCAAAT 2186
 DB AT 8505
 OY 2187 TTTTATATG---GAACAAGATTAAGATGAAAGTATTTGTTGGAAG 2242
 DB AT 8565
 OY 2243 TATTAATATACATTAAGCAAAATTTAAAAATTAATTAAGCCATTA 2302
 DB GTATATCTAGAGTATATTAATGATGCGATTTAGCAATTTGGCAACAAT 2362
 OY 2363 AACGAGTAAATTTTAACTTAAATTAATTAATTTGAGTAAATGTTTTC 2422
 DB AAAAAGATGATATTTTATTTATTAATAAATATATATTTTATTTTATAT 8745

OY 2423 TATTTAGGGGCAAAAAAGACAT 2447
 DB 8746 GTATATTAATTAAGATATATATAT 8770

RESULT 4

ABK28257/c
 ID ABK28257 standard; DNA; 8170 BP.

AC ABK28257;

XX 23-APR-2002 (first entry)

DE DNA transcription associated genomic DNA #66.

XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
 KW immunological disorder; Werner syndrome; developmental disorder;
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
 KW angiodysplastic disorder; congenital heart disease; HDR syndrome; gene therapy;
 KW polyglutamine disorder; solid tumour.

XX Unidentified.

XX WO200192565-A2.

PD 06-DEC-2001.

XX 06-APR-2001; 2001WO-EP03973.

XX 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI: 2002-090046/12.

XX New nucleic acids or oligomers, useful for diagnosing or treating

XX diseases associated with DNA transcription, e.g. immunological

XX disorders, Werner syndrome, psoriasis, myocardial infarction, solid

XX tumours or cancer

XX Claim 1; SEQ ID No 131; 32pp; English.

XX The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenosine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABK28117-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention.
 CC Note: The sequence data for this patent did not form part of the printed

OY 1937 ACAAAATTTTGAATAATGTTAAATTCCTAACATATATATAAATATGATGCTATA 1996
 DB 488 TATTTTGAATTAATTAATGTTAGTAAATTTTGTGAAATTTTGTGAAAT 547
 OY 1997 TGTATTTCCATATGTTCTTAATAATTTTGTATTTATTTAGTTATTAATATGATGAC 2056
 DB 548 TTTATTTATTAATGTTATTTGTTATTTAGTTATTAATTTATTTATTTATTTTAA 607
 OY 2057 CAATATAGTGTGCAATTCATAATCTCATTAATTTTGTGAATCTACAAATAT 2116
 DB 608 TATGATTTTGTATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTTAA 665
 OY 2117 AATATTTAGTCAATATACATGATGATGATGATGATGATGATGATGATGATGAT 2176
 DB 666 AATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 723
 OY 2177 CCAATTTTGTATTTTGTATGATGATGATGATGATGATGATGATGATGATGAT 2236
 DB 724 GATTTATTTATGTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTT 783
 OY 2237 TGGAAATGATTAATATACATTAATGATGATGATGATGATGATGATGATGATGAT 2296
 DB 784 ATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 838
 OY 2297 AATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 898
 DB 839 AATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 898
 OY 2357 AATGAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2416
 DB 899 GAAAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2416
 OY 2417 TCTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 958
 DB 959 GATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 996

RESULT 6
 ABL32426
 ID ABL32426 standard; DNA; 9810 BP.

AC ABL32426;
 XX

DT 26-MAR-2002 (first entry)
 XX

DE Human immune system associated gene SEQ ID NO: 399.
 XX

KW Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antiasthmatic; anti-HIV; anticonvulsant; ophthalmologic;
 KW antileukemic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.
 XX

PN WO200200928-A2.
 XX

PD 03-JAN-2002.
 XX

PF 02-JUL-2001; 2001WO-EP07537.
 XX

PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX

PA (EPIC-) EPIGENOMICS AG.
 XX

PI Olek A, Piepenbrock C, Berlin K;
 XX

DR WPI: 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PS Claim 1: SEQ ID NO 399; 32pp + Sequence Listing; German.
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 SQ Sequence 9810 BP; 2770 A; 79 C; 1978 G; 4983 T; 0 other;
 Query Match 2.68; Score 82.6; DB 24; Length 9810;
 Best Local Similarity 46.98; Pred. No. 0.00031;
 Matches 374; Conservative 0; Mismatches 404; Indels 19; Gaps 3;
 OY 1625 TAGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1684
 DB 822 TTTGTTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 881
 OY 1685 ACGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1741
 DB 882 GTGAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 941
 OY 1742 TGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1801
 DB 942 TTTTATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1001
 OY 1802 ATTAACGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1861
 DB 1002 AATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1061
 OY 1862 CTGCAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1921
 DB 1062 TATATATTAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 1121
 OY 1922 CATTTATTTAAGAAAAATTTTGAATA-----ANGTAAATTTCT 1966
 DB 1122 TTTTATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1181
 OY 1967 AACATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2026
 DB 1182 ATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1241
 OY 2027 TTTATATTTAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2086
 DB 1242 ATTAATTAATTTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1301
 OY 2087 ATTAATTTTGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2146
 DB 1302 ATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1361
 OY 2147 TTTCAAAAAATTTTGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2205
 DB 1362 ATTAATTAATTTGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1421
 OY 2206 ATTAAGATTAAGAAATTTTGTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2265
 DB 1422 AATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1481
 OY 2266 TTAATAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2325
 DB 1482 ATGATTAATTAATTAATTTGTTAAAGATTAATTAATTTATTAATTAATTAATTAATTAAT 1541
 OY 2326 AATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2385
 DB 1542 GTTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTATTAATTTATTTTATTTT 1601

[illegible]

Db	6832	AAATATAAAAACCAAAAAAATAAAAAATCAAAACAATTAATTAATTAATTTTCTACTA	6773
Qy	1460	TCTAAATATGCATGATCTTTTGGGTTACAGTTCCGAAGTCTCTACAAAGCGTGAACCAT	1519
Db	6772	TCTTAAATTAATTAATAAACCAAAATCAAAATAAAAAACAATATATTAACAATCT	6713
Qy	1520	CTGCACATATTAATTTGCTTCTTTATATGATCATCTTTACATATTAATTTATGTAGTGAAT	1579
Db	6712	CTTAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6633
Qy	1580	TTAATTAAGAGCGAATCTGTACATTAACAATATTTATATTAAGATACATGATGTGATATTT	1639
Db	6652	TATCTTTAAAAAATATAATTAATTTTAAAAAATAAAAAATTAACA - AAAAAAAAAA	6595
Qy	1640	CCAAATACATCTTTGGAGCTTTAACTTAATCTTGTTCTTCCTCCACGGTATTAATATTA	1699
Db	6594	AAAAAACAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6535
Qy	1700	ATCATGAGGATAAAAAAGTTTGTCTATTTTCGGATCATGAAGGATTAACCTATAG	1759
Db	6534	AATATATACCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6475
Qy	1760	ACTTATATTTTGGAAAAATGTAACCTTTTACTCATAGATTAATTAATTAATTAATTAATTAAT	1819
Db	6474	TTTCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6415
Qy	1820	TGCGAATATGACAGCCTCTACAACTGTGATGATCAATTTTTCGCAATATTAATTAATTAAT	1879
Db	6414	TAAATTCACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6355
Qy	1880	GAATTCAAATGCT---ACTATCAATTAAGAAACAGCTGATGATTAATTAATTAATTAATTAAT	1935
Db	6354	AAAAAAGCTTTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6295
Qy	1936	GACAAATTTTGGAAAAATGTAATTTCTTAACAATATTAATTAATTAATTAATTAATTAATTAAT	1995
Db	6294	TGATATATTTTCACTACATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6235
Qy	1996	ATGTAATTCCTATGTTCTTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2055
Db	6234	ATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6175
Qy	2056	CCAAATATATGTTGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2115
Db	6174	CTCTCTAAACACACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6115
Qy	2116	TATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2175
Db	6114	ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	6055
Qy	2176	TCCAAATTTTATTTTATGGAACAAGAAATTAACAGATTAAGAAATTAATTTGTGTGGA	2235
Db	6054	TCAACCATCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5995
Qy	2236	ATGGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2295
Db	5994	ATTTTTCAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5935
Qy	2296	AAAGATGTTATCTAGTAGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2355
Db	5934	AAATTCACACGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5875
Qy	2356	CAATGAACACGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2415
Db	5874	TATTAACAAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5815
Qy	2416	TTCTGACATATGA 2428	
Db	5814	CCCTATCTTTTCA 5802	

RESULT 9
 AAS45384/c
 ID AAS45384 standard; DNA: 5689 BP.
 XX
 AC AAS45384;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Chemically pretreated genomic DNA associated with cell cycle #45.
 XX
 KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KW graft-versus-host disease; glomerular disease; lewy body disease; cancer;
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritis;
 KW immunosuppressive; antitumour; cyostatic; antiarteriosclerotic; ds;
 KW PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN W0200168911-A2.
 XX
 PD 20-SEP-2001.
 XX
 PE 15-MAR-2001; 2001WO-EP02945.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIDENOMICS AG.
 XX
 PI Olek A, Plepenbrock C, Berlin K;
 XX
 DR WPI: 2001-602751/68.
 XX
 PT Designing primers and probes for analysing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle -
 XX
 PS Claim 1; SEQ ID NO 89; 28pp; English.
 XX
 CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers.
 CC
 XX
 SO Sequence 5689 BP; 1490 A; 207 C; 1277 G; 2707 T; 8 other;
 XX
 Query Match 2.58; Score 81.6; DB 22; Length 5689;
 Best Local Similarity 49.18; Pred. No. 0.00044;
 Matches 285; Conservative 0; Mismatches 292; Indels 3; Gaps 3;
 XX
 QY 1824 ATAAATGACAGCTCTACAACTGTATGTCATATTTTCTCCCAAAATTAATTAAGCAAT 1883
 DB 2110 AAAATTCGAAATCCGCAACCGCAATATCAAAAAATTTCTCCGAAAAAATAA 2051
 QY 1884 TCAATGCTACTATCAATTAAGAAACAGCTGATTAATTTAAATTAAGCAAAAT 1943
 DB 2050 TAT 1991

QY 1944 TTTTGAAAAATGTATTAATTTCTAACAAATTAATTAATAATGATGCTATATGTAATTT 2003
 DB 1990 ATTATTTAAAAATAATTTAAAAAANANATAAAAAATTTATTAANNNTAAAAAANNNTAN 1931
 QY 2004 CCTATGTTCTTAAATATTTTTTTTATATATAGTATTAATTAATTAATTAATTAAT 2063
 DB 1930 NATTAATAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1871
 QY 2064 AGTTGCT-GAATTCAAATATCTCCATTAATTTTGTGAAATCTCAATTAATTAATTAAT 2122
 DB 1870 AATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1811
 QY 2123 TAGTCATTAACATGCAAT-AGAAAGTTCCAAAAAATTTGTTAAACAGAACTTCCAAA 2181
 DB 1810 TATATAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1751
 QY 2182 TTTTTTTTTTTATGCAACAAGAAATTAACAGATACAAACATTTTGTGCGAATGGAA 2241
 DB 1750 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1691
 QY 2242 GTAGTAATATACATTAACCAATTTTAAAAAATTTATATAGCTTATACGCTCAAGTA 2301
 DB 1690 TAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1631
 QY 2302 TGTATCTAGTAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2360
 DB 1630 TATATCTATCATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1571
 QY 2361 AAAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2400
 DB 1570 ATTAATAAATAATTTTAAATACGTTAAATAATTAATTAATTAATTAATTAATTAAT 1531
 XX
 RESULT 10
 AAS46426/c
 ID AAS46426 standard; DNA: 5689 BP.
 XX
 AC AAS46426;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #148.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cyostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PE 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIDENOMICS AG.
 XX
 PI Olek A, Plepenbrock C, Berlin K;
 XX
 DR WPI: 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer -
 XX

PS Claim 1; SEQ ID No 148; 27pp; English.

CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (SS) and sequences complementary to (SS). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes. Sequences with even numbered Seq ID numbers are the
CC complementary sequence of the corresponding odd numbered sequence (e.g.
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5689 BP; 1490 A; 207 C; 1277 G; 2707 T; 8 other;

Query Match 2.5%; Score 81.6; DB 22; Length 5689;
Best Local Similarity 49.1%; Pred. No. 0.00044;

Matches 285; Conservative 0; Mismatches 292; Indels 3; Gaps 3;

DB 1824 ATATGACAGCCCTCTACACCTGATGATGCAATTTTTCGCAATATTAATTAAGCAAT 1883
DB 2110 AAAATGCCAATCCGCAAAACCGCAATATCAAAAATTTCCGCCGAAAAA 2051
QY 1884 TCAATGCTACTATCATAGAGAAACAGCTGAGATTTACATTTTAATTAAGCAAAAT 1943
DB 2050 TATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1991
QY 1944 TTTTGAATAATGTTAATTTCTACAAATTTAATTAATATGCTAATTTATTT 2003
DB 1990 ATTATTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1931
QY 2004 CCGATGTTCTAATAATTTTATTTATTTGTTAATAATACATTAACCAATAT 2063
DB 1930 NATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1871
QY 2064 AGTTGGT-GAATTCATAATATCTCCATTAAATTTTGAATAAGTACAAATTAATAT 2122
DB 1870 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1811
QY 2123 TATCATAATACATGCAT-AGAAAGTTCAAAAAATTTGTTAACAAGAACTTCCAA 2181
DB 1810 TATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1751
QY 2182 TTTTATTTTATTTTGAACAAGAAATTAACAGATAGAAACTATTTTGTGTGATGAA 2241
DB 1750 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1691
QY 2242 GTATGTAATATACATTAAGCAATTTTAAATAATATAGCTATAGCGCTCAAGTA 2301
DB 1690 TAACAATAAATAAATACGCAATATATTAATTAATTAATAAACAACCAAAAAACA 1631
QY 2302 TGTATCTGATGCTGATTAATA-ATGCATGCGCGGTTAGAAATGGGCAACAATG 2360
DB 1630 TATTAATCTATCAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1571
QY 2361 AAAACGAATTAATAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2400

DB 1570 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1531

RESULT 11

ID ABK28226/c

ABK28226 standard; DNA; 5689 BP.

ABK28226;

23-APR-2002 (first entry)

DNA transcription associated complementary genomic DNA #50.

CC PNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
CC DNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
CC single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
CC viral infection; Sezary syndrome; haematological disorder; tuberculosis;
CC immunological disorder; Werner syndrome; developmental disorder;
CC psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
CC neurodegenerative disorder; Mardenburg syndrome; Niemann-Pick disease;
CC myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
CC angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
CC polyglutamine disorder; solid tumour.

Unidentified.

WO200192565-A2.

06-DEC-2001.

06-APR-2001; 2001WO-EP03973.

06-APR-2000; 2000DE-1019058.

07-APR-2000; 2000DE-1019173.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIC-) EPIGENOMICS AG.

Olek A, Plepenbrock C, Berlin K;

WPI; 2002-090046/12.

Claim 1; SEQ ID No 100; 32pp; English.

CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Mardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from the
CC European Patent Office.

Db 2944 TTTTGTAGTTTGAATATTTTGTATGTTTTTAAAGTTAGATATGTAATTT 3003
Qy 2177 CCAAAATTTTTTTTTTATGGAACAATAACAGATAGAAAATTTTGTGGAA 2236
Db 3004 TATTTTATTTGATTTATGATGATGATTAAGAAAGACGATTTTTTGTGTAA 3063
Qy 2237 TGGAGTACTATATATCATTAAGCAATTTTAAAAATTTATATACCTATACCGCTCA 2296
Db 3064 AGATTTGAAGAGATGTTTTTTTATTTTGGAAAGTTAAAAAGAAAGGCTAG 3123
Qy 2297 AAGTATGATATGAGTAGGAGGTAATTAATTAACATGTCGATGCAATGGGCAAC 2356
Db 3124 GTTATGAAATAGTTATATATTTATGTTGTGTTAGGATTTAAATTAATTAATA 3183
Qy 2357 AATGAAACGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2416
Db 3184 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3243
Qy 2417 TCT 2419
Db 3244 TTT 3246

RESULT 13
ABL32891
ID ABL32891 standard; DNA; 6131 BP.
XX
AC ABL32891;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 864.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiasthmatic; anti-HIV; anticonvulsant; ophthalmologic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
KW antineoplastic; antiasthmatic; antidiabetic; antipsychotic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
PS Claim 1; SEQ ID NO 864; 32pp + Sequence listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX
SQ Sequence 6131 BP; 1970 A; 39 C; 1013 G; 3109 T; 0 other:
Query Match 2.5%; Score 80.4; DB 24; Length 6131;
Best local similarity 47.7%; Pred. No. 0.00068;
Matches 437; Conservative 0; Mismatches 456; Indels 23; Gaps 6;
Qy 1527 TATTAATGCTCTTCTTAATGCACTTTAAACATATTTATTTAGTGGAAATTAATA 1586
Db 357 TTTTAAAGATGTTTATTTTAAAGATATATGTTATATTAATTAATTAAGAAAAATTTTA 416
Qy 1587 GACGCACTTGTAACATTAATTAATTTATTTAGATACATGATGATTTCCAAATA 1646
Db 417 TTGAATTAATTAATTTGTAATTAATTAATTTATTTATTTAGGTTTAAATTTATATA 476
Qy 1647 CATACCTTGAGTGTAAATTAATTAATTTATTTCTCTTCTTACGATTAATTAATTAAT 1706
Db 477 TTTTATTTGATTAATTAATTAATTTATTTTATTAAGAAATTTTAAAGTTATTAAT 536
Qy 1707 AGGTAAAAAAGTTTGTCTTATTTTTCGATGATGATGATGATGATGATGATGAT 1766
Db 537 TAGTTGAAATGAATTAATGAATTTT---ATTAGGAAGATTAATTAATTAATTTGAA 592
Qy 1767 TTTTGAATTAATGAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1826
Db 593 TAAATTAATTAATTAATTAATTTATTTTATTTTATTTTATTTTATTTTATTTT 652
Qy 1827 ATGACAGCCCTCTCAACTGATGATGATGATGATGATGATGATGATGATGATGAT 1886
Db 653 ATGTTTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 705
Qy 1887 ATGCTACTATCAATAGAAAGAAACAGCTAGTATTAATTTTAAAGAAATTTT 1946
Db 706 ATTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 765
Qy 1947 TGAATAATGTTAATTTTCAACAATTTTAAATTAATTAATTAATTAATTAATTT 2006
Db 766 TTAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 819
Qy 2007 ATGCTTAAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 2066
Db 820 ATATTTTAAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 879
Qy 2067 TGGTAATTCAAATATCTCATTAATTTTGTGAATTCACAAATTAATTAATTTAGT 2126
Db 880 TGTAACTATAGAAATTTTAAAGATATTTT---AAATTAATTTTATTAATTTTGTGA 938
Qy 2127 CATTAACAATGATAGAAAGTTCCAAAAATTTGTTAAAGAACTTCCAAATTTT 2186
Db 939 AAAGTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 998
Qy 2187 TTTTATGGAACAAGAA---ATAACAGTGAAGAACTATTTGTTGGAATGGAAG 2242
Db 999 ATTTTGTATATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1058
Qy 2243 TAGTAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAAT 2302
Db 1059 TTTTGGATTTTAAAGAGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1117
Qy 2303 GTTATCTAGTAGGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2362
Db 1118 TTTTATTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1177
Qy 2363 AACGAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 2422
Db 1178 GATCTTTCGTTTATTAATTAATTTTGTAAAGTAAATTAATTAATTAATTAATTA 1237
Qy 2423 TATTGAGGGCAAAA 2438
Db 1238 TTTAAGAAATGTTGAA 1253

RESULT 14

ABA92787
ID ABA92787 standard; DNA: 640681 BP.
XX ABA92787;
AC ABA92787;
XX 27-MAR-2002 (first entry)
DE Buchnera sp. genomic DNA SEQ ID NO:1.
XX Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
KW circular; ds.
OS Buchnera sp.
PN JP2001292771-A.
PD 23-OCT-2001.
PE 07-APR-2000; 2000JP-0107160.
PR 07-APR-2000; 2000JP-0107160.
XX (RIKA) RIKAGAKU KENKYUSHO.
PA WPI: 2002-126043/17.
DR A genomic DNA of cockroach-symbiotic bacterium -
XX Claim 1; Page 16-230; 237pp; Japanese.
PS The present invention describes a gene (I) derived from Buchnera sp.
XX containing the DNA (a) or (b), (a) has a fully defined base pair
CC genomic DNA of ABA92787 given in the specification or is a DNA selected
CC from complementary DNA sequences, and (b) is a DNA which hybridises with
CC the DNA (a) and encodes a protein. Also described are: (1) a recombinant
CC vector (II) containing (I); (2) a transformant (III) containing (II);
CC (3) a genomic DNA of Buchnera sp. containing the sequence given in
CC ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
CC (d), (c) is a DNA containing a fully defined sequence given in ABA92788
CC or ABA92789 and (d) is a plasmid which hybridises with a DNA; and (5) a
CC method for the preparation of a protein in which (III) is cultured and
CC the expression protein of the objective protein is collected from the
CC resultant culture. The DNA is useful for developing agricultural
CC chemicals for exterminating cockroaches. The present sequence represents
CC the specifically claimed Buchnera sp. genomic DNA sequence, from the
CC present invention.
XX
XX
XX Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;
SO
Query Match 2.5%; Score 80; DB 24; Length 640681;
Best Local Similarity 45.8%; Pred. NO. 0.00086;
Matches 395; Conservative 0; Mismatches 455; Indels 12; Gaps 3;
QY 1559 ATATTATTGTTAGTGAATTTATAGAGGAGCTGTACATTACATTTATATTT 1618
DB 324686 ATATTGTTAGTGAATTTATAGAGGAGCTGTACATTACATTTATATTTGCAAC 324745
QY 1619 AGATACAGTATGCTATTTCCAAATACATCTTGGATGTTAACTTATCTGTTT 1678
DB 324746 AATCTTTTGTACAAAGAAATATATTTTTCATCATTAATCAATTTTGATC 324805
QY 1679 CTTCCTACGATTAATATTAATCATCGAGTAAAGATTTTGTCTTATTTTCGCAT 1738
DB 324806 TCTCTATCATGTAATAAATCAATTTCTTCTTAATATCAATTTTATATAT 324865
QY 1739 GCATCAAGGATAAAGCTATGCTTATTTTGAATGTAACCTTTACTATAGA 1798
DB 324866 TGTGGCTGATTTATGCTCATTAATAAATTTTAAAGATTTTCTTCT 324925
QY 1799 TTAATACGATGTTTGTGCAATGACACCTCTCAACATGATAGTCAATTT 1858
DB 324926 GTCAAGAAATATTTATTTTTCGTAAGAAATGATTAATGACAAATACGATATTT 324985

QY 1859 TTTCGCAATATTAATTAGAATTCATGCTACTATCAATAGAGAAAGCTGAGTA 1918
DB 324986 TTTTTCGATTAATTTACATTAATAATACATCTAGTAATTTTGANATTTAAATTAATA 325045
QY 1919 TTTCATTTTAAATTAAGCAAAATTTTGAAGAAATGTTATTTCTTACATATTTAT 1978
DB 325046 ATTTTTCGTTTAAATTTTAAATTTTGAATTTTCTATCATAGCAAGTATTTCTTAAT 325105
QY 1979 AAAATATGATGCTTAAATGATTTCCTA-----TCTCTTAAATATTTTATATAT 2033
DB 325106 AAGATTAATTTTATTTTAAATTTTAAATTTTAAATGCTATTTTAAATTTTATATA 325165
QY 2034 TTAGTTAATAATACATTTAAGCAATTAATAGTGGTGAATTCATATCTCCATTAATA 2093
DB 325166 ATTTAAATATTTCTTAAATTAATTTTAAATGCTATTTTAAATTTTAAATTAATA 325225
QY 2094 TTTTTCGAAATCTACAAAT-----TATTAATTTTATGTAATTAACAAATGCAATTAAGTT 2148
DB 325226 ATTTATTTTATTTTACATTTTATTAATTAATTAATTAATTAATTAATTTTATTA 325285
QY 2149 CCAAAAAAATTTTGTAAACAGAACTTCCAAATTTTATTTTATGGAACAGAAATA 2208
DB 325286 ATTAATAATACATTTTATTAATTAATTAATTAATTAATTTTCTTTTATTTTAAATTT 325345
QY 2209 ACAGATAGAAATCTATTTTGTGGAATGAGATTAATTAATTAATTAATTTTAAATTTTAA 2268
DB 325346 ACTATTAATGATA-ATTATAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 325403
QY 2269 AAAATTTATTAAGCTATACGCGCTCAAGATGATGTATCTAGTGTATTAATTAAT 2328
DB 325404 TCAATTAATAATTAATTAAGATTAATTTCTTTTATTTTATTAATTTTATTAATTA 325463
QY 2329 GCATGTCGATTCAGAAATTTGGGACACAAATGAAACGAATTAATAATTAATTAATTA 2388
DB 325464 GAATTTGCATGCAATCGAATTAATTAATTTTATTTTCAATGATTAATTAATTAACAA 325523
QY 2389 ATTAATTAATAATTTGATTAAT 2410
DB 325524 TCAATTAATAAATCTTATTAATTT 325545
RESULT 15
ABK31511/c
ID ABK31511 standard; DNA: 47108 BP.
XX ABK31511;
AC
XX
XX 23-APR-2002 (first entry)
DE
XX
XX Signal transduction associated gene modified complementary DNA #177.
XX Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytosine; mutant; ds.
OS Homo sapiens.
OS Synthetic.
PN WO200200926-A2.
PD 03-JAN-2002.
PE 29-JUN-2001; 2001WO-EP07472.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Phippenbrock C, Berlin K;
PI WPI: 2002-147896/19.
DR

XX Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction
XX
XX Claim 1: SEQ ID No 354; 24bp; English.

CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (CpG islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK1158-ABK11545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
XX European Patent Office.

XX Sequence 47108 BP; 15349 A; 429 C; 8736 G; 22594 T; 0 other;

Query Match

Best Local Similarity 2.58; Score 79.6; DB 24; Length 47108;
Matches 457; Conservative 0; Mismatches 564; Indels 9; Gaps 3;

QY 1074 TTCAGCTATTTTACAAAGTACAAAGTATATGCTTGAATCCCTTGAATAATTTT 1133
DB 1033 TTAATTCCTTTACCAAAAAACGATTTATATTAATTAACCAAACTTAATAATCT 974
QY 1134 GGAATTCGGTGTGTATGTAATATAGTTAGTGTAAACAATAATGTAATCAATTA 1193
DB 973 AAACCTTAAAAATTAATAATTAATCAAAATTCATTTCTTAATTTAATTAATTAATA 914
QY 1194 GTGGTCAACATATACATATATCTTACGAGAAAAACAACCTTAAGAGAGTAAATATC 1253
DB 913 ATTAATTAATAATATCTTCTTATTAATAAATCAATTAATAATTAATTAATTAATA 914
QY 1254 CATATATGGTATGCTATACCTTTCAGGTATGCTATAGAGACTAA---AGATAGTT 1310
DB 853 TATTCCTGAATATTTAAAAAATAATATATATCTTACAGCTCTCTATTAATAATTA 794
QY 1311 ATGTGATGTGATTAATGAATTCACACGGGTGTAATAATTAATGAGACCTATGTTAG 1370
DB 793 AAATTAATTAATAATTAACACATACACACTAATCTAAATAATTAATCAATTAATAC- 735
QY 1371 ATCACTCAATATATCATCTTGTGGTGTGCAATTAATAAACAATAAAGAAAAAGAA 1430
DB 734 TTTCTTAAATTAACACATATATCAATTAATCTTAAACATTAATCTTCTCAAAAA 675
QY 1431 AACGATTTTCTGATTCATTCATGATCTTAATAATGATAGTCTTTGGTTCAGT 1490
DB 674 AAATCCATTAATAAACAATTAATTTCTTAATAATTAATTAATTTCTTAATAATACAT 615
QY 1491 TTGGAAGTCTCTACAGGTGTAAACCATCTGCACTATTAATATGCTTTTAATGCA 1550
DB 614 TCCCAATTAACCAACACTTTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 555
QY 1551 TCTTAACATATTAATGTTAGTGAATTAATAAGAGCACTGTAACTATACATA 1610
DB 554 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 495
QY 1611 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1665
DB 494 TATAACTCCAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 435

QY 1666 CTTATATCTGTTCTTCTCAGGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1725
DB 434 TTTACCTTTTAACCTTTTAAAAATTTTACCTTATTTTATTTTCTCTCAATATCTAATTT 375
QY 1726 TTTATTTGCGATGATGATGAGATTAACCTAATGATCTTAATTTTAAAAATGTAACC 1785
DB 374 TCAATCTCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 315
QY 1786 TTTTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1845
DB 314 TTTTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 255
QY 1846 TGAATGCAATTTTCTGCAAAATATTAATTAATTAATTAATTAATTAATTAATTA 1905
DB 254 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 195
QY 1906 AAACGCTGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1965
DB 194 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 135
QY 1966 TTAACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2025
DB 134 CTAATAATTAATAAATTTCTATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 75
QY 2026 TTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2085
DB 74 CTAATAATTAATAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 15
QY 2086 CATTAAATTT 2095
DB 14 TTTTAATTT 5

Search completed: March 29, 2003, 18:59:11
Job time: 2214.58 secs

Handwritten scribbles and marks, possibly a signature or initials.


```

; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
;
US-08-998-416-288

```

Query Match	2.1%;	Score 68.6;	DB 4;	Length 837;
Best Local Similarity	46.9%;	Pred. No. 4.6e-05;		
Matches 250; Conservative	0;	Mismatches 279;	Indels 4;	Gaps 1

[illegible]

RESULT 3
 US-08-487-826B-13/C
 : Sequence 13: Application US/08487826B
 : Patent No. 593827
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Slim, Kam L.
 : APPLICANT: Chitnis, Chetan
 : APPLICANT: Miller, Louis H.
 : APPLICANT: Peterson, David S.
 : APPLICANT: Su, Xin-zhaun
 : APPLICANT: Wellens, Thomas E.
 :
 : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 : NUMBER OF INVENTIONS: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 : NUMBER OF SEQUENCES: 45
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: Knobbe Martens Olson & Bear
 : STREET: 620 Newport Center Drive 16th Floor
 : City: Newport Beach
 : STATE: California
 : COUNTRY: US
 :
 : ZIP: 92660
 :
 : COMPUTER READABLE FORM:
 :
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 :
 : CURRENT APPLICATION DATA:

1 APPLICATION NUMBER: US/08/487,826B
2
3 FILING DATE: 10-SEP-1993
4
5 CLASSIFICATION: 435
6
7 ATTORNEY/AGENT INFORMATION:
8
9 NAME: Israelsep, Ned
10
11 REGISTRATION NUMBER: 29,655
12
13 REFERENCE/DOCKET NUMBER: NH121.001CIP
14
15 TELECOMMUNICATION INFORMATION:
16
17 TELEPHONE: (619) 235-8550
18
19 TELEFAX: (619) 235-0176
20
21 INFORMATION FOR SEQ ID NO: 13:
22
23 SEQUENCE CHARACTERISTICS:
24
25 LENGTH: 19124 base pairs
26
27 TYPE: nucleic acid
28
29 STRANDEDNESS: single
30
31 TOPOLOGY: linear
32
33 MOLECULE TYPE: DNA
34
35 HYPOTHETICAL: NO
36
37 ANTI-SENSE: NO
38
39 US-08-487-826B-13

Query Match	2.1%	Score 67.4;	DB 2;	Length 19124;
Best Local Similarity	42.3%;	Pred. No. 0.00016;		
Matches 821; Conservative	0;	Mismatches 1091;	Indels 29;	Gaps 7

[illegible]

PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: 1123..3123
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3871..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5880
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5996..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon
LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc-feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTGT
NAME/KEY: allele
LOCATION: 2048

OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G

QY	1842	ACIGGATAGCAATTTTTTCTGC	CAAAATTTAAATTAGAATTC	CAATGCTACTATCAATA	1901
Db	11120	AATATATAAATTTAAATTTTAA	TAAATTTAAATTTAAATTTTA	TAAATTTAAATTTTA	1178
QY	1902	GAGCAACAGCGTAGATTTACAT	TTTAAATTTAAAGACAAATTT	TGAAAAATGTTATAA	1961
Db	11179	ATTAAATTTAAATTTAAATTTAA	ATTTAAATTTAAATTTAAATTT	AAATTTAAATTTAAATTTAA	11238
QY	1962	TTTCTAACAAATTTTAAATTA	AAATTAAGATGCCATTAATG	TATTTCCATGTTCTTAAATAT	2021
Db	11239	TTTTTAAATTTAAATTTAAATTT	AAATTTAAATTTAAATTTAA	CTTAATTTAAATTTAA	11298
QY	2022	TTTTTTTAAATTTTGGTAAAT	TAAATTCATTAAGAACCAATTA	ATAGTTGGGATTCAAATA	2081
Db	11299	TATTTATTTAACTTAAT-TA	TTTAAATTTAAATTTAAATTT	TAATTTAAATTTTAAAT	11357
QY	2082	TCTCCATTAAT--ATT	TTTTGGAATCTCAAAATTT	TAATTTTAAATTTAGTCAATTA	CAATGC 2138
Db	11358	TTAATTAATTTAAATTTAAATTT	AAATTTAAATTTAAATTTAA	TTAAATTTAAATTTAAATTT	ATG 11417
QY	2139	ATGAAACTTCCAAAAAAT-T	TGTTAACGAACCTTCCAAAT	TTTTTTTTTTTATG	2197
Db	11418	CTCAATTTAAATTTAAATTTTAA	ATTTTAAATTTTAAATTTAA	ATTTAAATTTAAATTTTAAAT	TTT 11477
QY	2198	AACAGAAGATAACAGATGA	AAACATTTTGGTGCAATGGA	GTAGATATATACATTA	2257
Db	11478	AAAATTAATTTAAATTTAAATTT	AAATTTAAATTTAAATTTAA	ATTTAAATTTAAATTTAAATTT	TAAGCTAAG 11537
QY	2258	AGCAAAATTTAAAAATTTAT	ATATAGCC 2284		
Db	11538	AAAAATTTTATATTTAAAAA	AGGCC 11564		

Our

Query Match	Similarity	2.0%	Score 63.6	DB 2	Length 8920
Best Local Similarity	46.8%	Pred. 0.00074			
Matches 278	Conservative	0	Mismatches 304	Indels 12	Gaps 2
QY	1874	AATTAGAAATTCAAATGCTACATATCAGAGAAACAGCTGATATTAACATTTTAATTA	1933		
Db	152	AAAAATATATATAAATATAAAAAATATAAACATATAAAATATATATTTGATATGTAAT	211		
QY	1934	AAGACAAATTTTGGAAAAATGTTATA-ATTTCTAACATATATTATTAATAATGATGCT	1992		
Db	212	ATATATGATATATTCATATTAATTAACATAGATAAAAAACTTTTTTTTTTTTTTCT	271		
QY	1993	ATAATGATATTCCTATGCTCTTAAAAATTTTTTTTATATTAGTTATTAATACATTAT	2052		
Db	272	TATATTTATTAACATACATACATTAAGTTATTTTATATATATATATATATATATATA	331		
QY	2053	GAAACAATATATGTTGGGGAATTCAAATATCTCCATTAATTTTTTGAACCTACAAAT	2112		
Db	332	TATATATATATATATGTTGGTGGTTCATTTGTTATATAAAATCTGAAATATAAAGT	391		
QY	2113	TATTAATATTTAGTCAATTAACATGATCAGAAAGTTCCAAAAAAATTTTGTATACAGA	2172		
Db	392	TATTAATATATTTCCAAATTAATATGAAATBCAATTTATTAATATTTGATGCTGACACATTA	451		
QY	2173	ACTTCCAAATTTTTTTTTTTATGGA-----ACAAAGAAATTAACAGATAGAAAC	2221		
Db	452	ATATAGTTTATACACTCTTATATAAACCATCGTATATATATACACAATATATATAC	511		
QY	2222	TATTTGTTGGGAATGGAAGTAGTAAATACATTTAAGCAAAATTTAAAAAATATATATA	2281		

D_b 512 TCCCAATATGTGTCCTAATAATTTAAATTAATTAATTAATTAATTAATGCATT 571
O_y 2282 GCCATACGCCCTAAAGTATGTTATCTACTAGCTGAATTAAATGATGGCGATT 2344
D_b 572 ATTTATTTTTTTCTTGTTATTAATAATCTATCTACTAATTTAAAAA
O_y 2342 CAGAATTGGGACAACATGAAAACGCAATTAATTTAACCTTAAATAATAAATNT 2401
D_b 632 AAAAAAAAAAAAAAAAAAAAAAAAAATTCATTAAGAAAAATGAACTTGTAATGTA 691
O_y 2402 TGAGTAATATGTTTCTGCATTTGAGGGGCAAAAAAGCAATGCACAAAG 2455
D_b 692 TTATTAATATTTTAAACATTAATATTAATGCTTAAAAAAGAAAAATG 745

```

RESULT 8
US-09-150-741-1
: Sequence 1, Application US/09150741
: Patent No. 6183996
: GENERAL INFORMATION:
: APPLICANT: Stewart et al.
: TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
: Patent No. 6183996
: TITLE OF INVENTION: Synthetase II
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/09/150,741
: CURRENT FILING DATE: 1998-09-10
: EARLIER APPLICATION NUMBER: PLS380
: EARLIER FILING DATE: 1992-12-16
: EARLIER APPLICATION NUMBER: A093/00617
: EARLIER FILING DATE: 1993-12-02
: EARLIER APPLICATION NUMBER: 08/446,855
: EARLIER FILING DATE: 1995-07-06
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 1
: LENGTH: 8920
: TYPE: DNA
: ORGANISM: Plasmodium falciparum
US-09-150-741-1

```

Query Match	2.08;	Score 63.6;	DB 4;	Length 8920;
Best Local Similarity	46.88;	Pred. NO. 0.00074;		
Matches 278; Conservative	0;	Mismatches 304;	Indels 12;	Gaps 2

QY	1874	AATTAGCAATTCATGCTACTACTACATTAAGAGAACAGCTGAGTATTTACATTTTAAATTA	1933
Db	152	AAAAAAAAATATATAATTAATAAATAAAGATAAAAAATTTTATATTTGATATGATAGTAT	211
QY	1934	AAGCAAAAATTTTTGGAAAAATGTTATA-ATTCTTAACAATATATTTAAATATGATGCT	1992
Db	212	ATATAAAGATTAATTCATATTAATTAATACATAGATAAAAAACTTTTTTTTTTTTTTTCT	271
QY	1993	ATATATGATTTCCATGCTCTTAAATATTTTTTTTTTATATTTAGTATTAATACATATAT	2055
Db	272	TTATATTTTATTAACAATACATATTTAAGTTATTTTATATATATATATATATATATATATA	331
QY	2053	GAACCAATTAATAGTGGTGGATTCACAAATATCCCATTAATATTTTGGAAATCCTCAAT	2112
Db	332	TATATATATATATATATGTTCTGTCTTCATTTGTTTATTAATAATTTACTGAAATATATAA	391
QY	2113	TATTAATATATTTAGTCATAAACAATGCAATGATAGAAAGTTCCAAAAAAAATTTTGTAACAGA	2172
Db	392	TATTAATATATTTCCATTAATATATGAAATCAAAATATATATATTTTGATGTGACACATTA	451
QY	2173	ACTTCCAAATTTTTTTTTTATGGA-----ACAAGAAATTAACAGTAGAAAAAC	2221
Db	452	ATATAGTTTACACTCTCTTAAATAAACACATCCATATATTTATACACATATATTAATAC	511
QY	2222	TATTTGTGGATGAGTAGATATATACATTAAGCAAAATTTTAAAAAATTTATTA	2281
Db	512	TCCCAATATATGGTGTCCATATATTTTATTTATATATTTATTTATTAATTTATATCAATTT	571

OY 2282 GCGCATACGGGCATCAAGTGTGTTACTCAGTCGTATTAATTATATGATGCTGCGAATT 2341
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dd 572 ATTTATTTTTTCTTGTTAGTTATAAAATAGTAATCTACTAAATTTAAAAAAAAAAAAAAAA 631

OY 2342 CAGAATTGGGACAACATGAAGAACGCAATTAATATTAATCTTAAAATTAATAAAATT 2401
|| | | | | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dd 632 AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAATTTACATATGCCAAAAATGAACTGTTATGTAA 691

OY 2402 TGAGTAATGTTGTTTCTGCAATTGNAGGGCAAAAAAAAAAGACAAATGCCAAAG 2455
- - - - - ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dd 692 TTTTATAATTTTTAAACATAAATATTAATATGTTAAAAAAAAAAAAAAAAAGAAAAATG 745

```

RESULT 9
US-09-641,638-651/C
Sequence 651, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Anick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET_05ICPI
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/215,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651

```

```

1 ORGANISM: Homo sapiens
2 FEATURE:
3 NAME/KEY: misc_feature
4 LOCATION: 1123..3123
5 OTHER INFORMATION: 5 regulatory region
6 NAME/KEY: exon
7 LOCATION: 3124..3297
8 OTHER INFORMATION: exon 1
9 NAME/KEY: exon
10 LOCATION: 3871..4072
11 OTHER INFORMATION: exon 2
12 NAME/KEY: exon
13 LOCATION: 5552..5633
14 OTHER INFORMATION: exon 3
15 NAME/KEY: exon
16 LOCATION: 5758..5880
17 OTHER INFORMATION: exon 4
18 NAME/KEY: exon
19 LOCATION: 5996..6099
20 OTHER INFORMATION: exon 5
21 NAME/KEY: exon
22 LOCATION: 6349..6509
23 OTHER INFORMATION: exon 6
24 NAME/KEY: exon
25 LOCATION: 7379..7522
26 OTHER INFORMATION: exon 7
27 NAME/KEY: exon
28 LOCATION: 8645..8854
29 OTHER INFORMATION: exon 8
30 NAME/KEY: exon
31 LOCATION: 12254..12340
32 OTHER INFORMATION: exon 9
33 NAME/KEY: exon

```


LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc_feature
LOCATION: 17555..20674
OTHER INFORMATION: 3' regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903

OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match 2.0%; Score 63; DB 4; Length 20674;
Best Local Similarity 48.0%; Pred. No. 0.0012;
Matches 224; Conservative 0; Mismatches 230; Indels 13; Gaps 1;

Db 4739 TACTAGTCAAGCTAATTAATTTCTATTTTAAAG 4779

RESULT 11

US-07-867-106-2/c

Sequence 2, Application US/07867106

Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B

APPLICANT: Chang, Andy C M

APPLICANT: Williams, Keith L

TITLE OF INVENTION: Improved Plasmid Vectors for Cellular

TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5389526r1s

STREET: One Liberty Place 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867,106

FILING DATE: 19920625

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU 87 7187

APPLICATION NUMBER: PCT/AU90/00530

FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Peeney, Joanne Longo

REGISTRATION NUMBER: 35,134

REFERENCE/DOCKET NUMBER: RICE-0002

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 5852 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

US-07-867-106-2

Query Match

Best Local Similarity 48.3%; Pred. No. 0.0015;

Matches 209; Conservative 0; Mismatches 217; Indels 7; Gaps 1;

1.9%; Score 61.8; DB 1; Length 5852;

Query Match

Best Local Similarity 48.3%; Pred. No. 0.0015;

Matches 209; Conservative 0; Mismatches 217; Indels 7; Gaps 1;

1.9%; Score 61.8; DB 1; Length 5852;

Query Match

Best Local Similarity 48.3%; Pred. No. 0.0015;

Matches 209; Conservative 0; Mismatches 217; Indels 7; Gaps 1;

1.9%; Score 61.8; DB 1; Length 5852;

Db 5607 AAAAAAATGTAATTAATTTGGAATTAATAAATGTAAGGCTTTT 5548

2007 ATGCTTAATAATTTTATA-----TTAGTATAATTCATTAAGACCA 2059

5547 AAAAAATGTAATTTTATAATTTTAAATCATTGACGAGATTAATAA 5488

2060 TAATAGTGTGTAATCAATATCTCATTAATTTTGAATTCATAATTAAT 2119

5487 TAAACATATTTGATTTTATTTTATTTTATTTTATTTTAAATCAATTAATA 5428

2120 ATTTAGTCAATTAATGCTAGAGAAAGTCCAAAAATTTGTTAAGCAACTTCA 2179

5427 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5368

2180 AATTTTATTTT 2192

5367 AATTTAGTATCT 5355

RESULT 12

US-09-316-083-2

Sequence 2, Application US/09316083A

Patent No. 6280942

GENERAL INFORMATION:

APPLICANT: The Institute of Physical and Chemical Research

TITLE OF INVENTION: Endonuclease

FILE REFERENCE: PH-651

CURRENT APPLICATION NUMBER: US/09/316,083A

CURRENT FILING DATE: 1999-05-20

EARLIER APPLICATION NUMBER: JP98/141861

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 2

LENGTH: 1431

TYPE: DNA

ORGANISM: Saccharomyces cerevisiae

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1428)

US-09-316-083-2

Query Match

Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

1.9%; Score 61.2; DB 4; Length 1431;

Query Match

Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

1.9%; Score 61.2; DB 4; Length 1431;

Query Match

Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

1.9%; Score 61.2; DB 4; Length 1431;

Query Match

Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

1.9%; Score 61.2; DB 4; Length 1431;

Query Match

Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

1.9%; Score 61.2; DB 4; Length 1431;

Query Match

Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

1.9%; Score 61.2; DB 4; Length 1431;

Query Match

Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

1.9%; Score 61.2; DB 4; Length 1431;

Query Match

Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;

1.9%; Score 61.2; DB 4; Length 1431;

Query Match

QY 1559 ATTTTATTTGTTAGTTGGAAATTTAATAGACGCACTGTAACTTAACAATATTTATATT 1618
Db 324666 AATATTTAGTATTCCTTTTTTATTTATTCACAAATATTTTATTCATCAATTAATTTGGCAAC 324745

QY 1619 AGATACTAGTATGTGATTTATTCAGAAATACACTTGGATGTTTAACTTAATCTGCTTT 1678
Db 324746 AATCTTTTGTACAAAGAAAAAATATATTTTTTTTATCTATATAATCCATTTTGTATC 324805

QY 1679 CTTCCTACGGTATTAATATTTATTCATCGAGGTAAAAAAGTTTGTATTTTCGGCAT 1738
Db 324806 TCTCTCTATCAAGTAAAAATATCATCTTTCTTTATATCAATTAATTTTAAATATAT 324865

QY 1739 GCATGAAGGATTAACCTAATGACTTTAATTTTTGAAATAGTAACCTTTACTCATAGA 1798
Db 324866 TGTGGGTGTATTTTATGTCTCATTAATAAAATCTTAATAATTTTTTTTAAAGATTTTCTTC 324925

[illegible]

Db	398	TAATATTTTTTCCCAATATATAATATCCCTCATTTTTTTCCTTTTTTATTCTTTAT	339
QY	68	TTCCACATGATTTGAGTTTGGTCTCAATTTGGATGCCAGATAAATTAATATAATTC	127
Db	338	TTTTTAATTTTAAATACCTTTTTTTTTTAAATAACTAAATCTTTTTTAAATTCAT	279
QY	128	ATTTAAATATTACAGATTAATTAATCTTTACATTGTATGTTATACAAAATATCT	187
Db	278	TTTTTTTTTTCTTAAAAATCTCTTTTTTAAATTTTTTAAATTTCTTAAATATTTTTT	219
QY	188	ATCTTGGTATATGAGAAAATATGAGTTTGGATTTTAAATAAAGAAATATACGA	247
Db	218	TTTTTAATTTTTTAAACAAATATTAAATTTAAATTTAAATTTTAAATTTTTTTTA	159
QY	248	TTTCATTTGGTGATTACACAGTTAGTTTGTGTCTTTGTTATATGATATAGAG	307
Db	158	TTTTTTTTTATTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTATTTTAAATTTAA	99
QY	308	TAATATCAAAAGAGTATGATTTGAAGTGTAAACATATTTCTTATGACCCCAAAAAA	367
Db	98	AA	39
QY	368	AAAAAAAAACAAACAACAA 387	
Db	38	AAAAAAAAAAAAAAAAAAAAA 19	

```

RESULT 3
US-09-754-853A-2
: Sequence 2, Application US/09754853A
: Publication NO. US20030005491A1
: GENERAL INFORMATION:
: APPLICANT: Haugé, Brian M.
: APPLICANT: Parnell, Laurence D.
: APPLICANT: Parsons, Jeremy D.
: APPLICANT: Wang, Ming Li
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: FILE REFERENCE: 38-10(15810)B
: CURRENT APPLICATION NUMBER: US/09/754,853A
: CURRENT FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 60/174,880
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 1119
: SEQ ID NO 2
: LENGTH: 335913
: TYPE: DNA
: ORGANISM: Glycine max
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (45163)..(45314),(45450)..(45509),(46941)..(48763),(48975)..(49573)
: OTHER INFORMATION: Clone ID: 24001_region_63
: US-09-754-853A-2

```

	Query Match	Best Local Similarity	46.1% ; Score 65.2 ; DB 9 ; Length 335913 ;
	Matches	399 ; Conservative	0 ; Mismatches 458 ; Indels 9 ; Gaps 5 ;
QY 1582	AATAAGACCGAAGCTGTACATTACATAATTTATATTTGATACACTAGTATGCGATTATGCC		1641
Db 279575	ATTATATATATAAAATATATAATTTATTAACATTTTAAAGAAAAATATATACAGTAAAAATTTAT		279634
QY 1642	AAATACACTACTCTTGATGTATTAACCTTAATCTGTCTTCCTACGTATATAATTTATAT		1701
Db 279635	TTTAAATCAAAAAATTCCTCTCTTTATTTTATTTTGTATTTATTTTAAAGACAAGATTAAACAT		279694
QY 1702	CATCGAGGTAAAAAAGTTTGCTCTATTTTCGCGATCAATGAAGATTAACCTATATGAC		1761
Db 279695	AACGATTTTCAAAATAGCTTAGCTTTGATTTTACGACAGTGAATVCGGTAAAGACCCA		279754
QY 1762	TTTAAATTTTGGAAAAATGAACCTTTTACATCATAGATTAATTTACCGTATGTTTTGTGTG		1821

Query Match	2.08;	Score 65.2;	DB 9;	Length 335913;
Best Local Similarity	46.18;	Pred. No. 0.67;		
Matches 399;	Conservative	0;	Mismatches 458;	Indels 9;
			Gaps 5;	
QY 1582	AATAAGACGCACTTGGACATCAATATTTATATAGTACTAGTATGTGTTTC	1641		
DB 279575	ATTATATATATAATATATATATATATATATATATATATATATATATATAT	279634		
QY 1642	AAATACATCTTGGATGTTTAAAGCTTAATCTGTTTCTTCTACGGTATTAATTAAT	1701		
DB 279635	TTTATATCAAAAATTCCTCTTTTATTTTATTTAGTATATATATATATATATATAT	279664		
QY 1702	CATCGAGTAAAAAAATTTTGTCTTATTTTGGCGATGCAATGAGATTAACCTAATGAC	1761		
DB 279695	AACGATTTCCAAAATACCTTACGCTTGATTTTCGACAGCTGGAAATCGGTAAGACCCA	279754		
QY 1762	TTTTATTTTGTGAAATGTATACCCCTTTTACTCATATATTAATTAATTAATTAATTAAT	1821		
DB 279755	ACATATTCCTT-TTGCCTTTTCAATTTTCAATTCATTCGCGGTTTCTTTTCAATTTATGTTTC	279813		
QY 1822	CCATTAATGACAGCCCTACAACTGATATGTCATTTTTCGCAATATTTATTAAGA	1881		
DB 279814	CTTAAT	279873		
QY 1882	ATTCATGCTACTATCAATATGAAAGAAACAGCTGATATATACATTTTAATTTAAAG-ACAA	1940		
DB 279874	TAAAGAAAGTATATGATATAAAAACCTAATTAACCTAATCCATATATTTTGATCCGAT	279933		
QY 1941	AATTTTGGAAAAATGTTATATATTTTCAACAATATATATATATATATATATATATATATAT	2000		
DB 279934	AAATATCGAATCAGATATATATATTTTCAACATGCTTAATTAATATATATATATATATATAT	279993		
QY 2001	TTTTCTATGTTCTTAAAAATATTTTATATATATATATATATATATATATATATATATATAT	2060		
DB 279994	AT	280049		
QY 2061	AATAGTTGGTAATTCAAATATCTGCATTAATATTTTGAATCTACAAATATATATA	2120		
DB 280050	ATTATATTTTAAACAATATATATGCAAAATGTATTTTATTTATGCAATTTCTTTCTTA	280109		
QY 2121	TTTATGCAATTAACAATGATATAGAAAGTCCAAAAATTTTGTATACAGAACTTCGAA	2180		
DB 280110	--AAATTAATAACAAT	280167		
QY 2181	ATTTTATTTTATATAGAACAAAGAAATATACGATATGAAAACTATTTTGTGTGAATGGA	2240		
DB 280168	ATTATTTATATTTTAAAAATTTTAAATTAACAATTTTATTTATTTTAAATTAATGATGAC	280227		
QY 2241	ACTAGTAATATATACATTAAGCAAAATTTTAAAAATTTATATATAGCCTATACGCGTCAAGT	2300		
DB 280228	AAACTAT	280287		
QY 2301	ATGTATCTATAGTGTGTAATTAATATATGCAATGCTGCGATTCAGAAATGGACACAAATG	2360		
DB 280288	ATTAGTAT	280346		
QY 2361	AAAAAGATTAATAAT	2420		
DB 280347	TTATATACATATATCTAT	280406		
QY 2421	ACTATATGAGGGCAAAAAAGACAA 2446			
DB 280407	TTTATATATCTGTTTAACTAATTCAGAA 280432			

```

; TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
; FILE REFERENCE: 38-10(52679)A
; CURRENT APPLICATION NUMBER: US/09/969,373
; PRIOR APPLICATION NUMBER: US 09/754,853
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 09/760,427
; PRIOR FILING DATE: 2001-01-13
; PRIOR APPLICATION NUMBER: US 09/855,768
; PRIOR FILING DATE: 2001-05-15
; NUMBER OF SEQ ID NOS: 4593
; SEQ ID NO 1212
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Glycine max
US-09-969-373-1212

Query Match      2.0%; Score 64.4; DB 10; Length 473;
Best Local Similarity 52.2%; Pred. No. 0.083;
Matches 166; Conservative 0; Mismatches 151; Indels 1; Gaps 1;

QY 1866 AATATTAAATTTAGGAATTCAGTCTACTATCAATAGAGAAGAACGCTGACTATTACATT 1935
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 431 AAATACCTAAATTAATTAATTTTACCTATTTCTAATTAATTTTATTTATTTATTTATTT 372

QY 1926 TTAATTAAACACAAATTTTGGAAAAATGTTATTAATTTCTAACAAATTTTAAATAT 1985
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 371 AAATCCCTTAATTAATTAATTTTATTTAGTCTGCTGCAATTTTTTA-ACCTTATATAATTC 313

QY 1986 GATGCCCTTAATGTATGTATGCTATGCTTAAATATTTTATATTTTATGTTATATAAT 2045
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 312 CAATTTTATGCTTTATTTCTCTAATTAATTAATTAATTTATTTATTTTTCGCTATTAC 253

QY 2046 ACATTATGACCAATTAATAGTTGGTGAATTCAAATATCTCCATTAAATATTTTGGAAATC 2105
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 252 AAATTAATTAATTTATTTATTTAGAAAAACAAACGTAATAATGTTAATTAATTAACACATAAAA 193

QY 2106 TACAAATTTATTAATTTTATGCAATTAACATGCAATAGAAAAATTCACAAAAATTTTGT 2165
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 192 TAAATTTATTTATTTTATTTAGACATTAAATTAATTAATTAATTAATTTATTAAGAAATTAATCTA 133

QY 2166 AACAGAAACTTCCAAAATT 2183
      ||||| ||||| |||||
DB 132 AAAATTAATATTTATTTT 115

RESULT 6
US-09-960-352-11234/C
; Sequence 11234, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11234
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-LIB3058-052-Q1-K1-D8
US-09-960-352-11234

Query Match      2.0%; Score 64.2; DB 10; Length 419;
Best Local Similarity 49.9%; Pred. No. 0.086;
Matches 189; Conservative 0; Mismatches 188; Indels 2; Gaps 1;

Y 1842 ACTGTGATAGTCGAATTTTCTCGCAAAATATTAAATTAGGAATTCATGCTACTATCATA 1901

```

```

Db      417 AATTTATTAATATATATTTTAAAAAATTAATTTTATATATTAATTTTATATA 358
Oy      1902 GAAGAAACAGCTGAGTATTCATTTTAATTTAAAGCAAAATTTTGGAAAAATGTTATTA 1961
Db      357 AAATATTTTAAAAAATTTTAAAAATTTTAAAAATTTTATATATTTATTTAAATATATA 298
Oy      1962 TTTCATACAAATTTATTTAAATATGATGCCCTATATGTATTTCCATATGTTCTTAATAT 2021
Db      297 TTTTAAAAAATTTTAAAAATTTAAT--TTAAATTTAAAAAATATATATATAAAAAAT 240
Oy      2022 TTTTATATATTTAGTTAATATCATATATACACCAAAATATAGTTGGTAATTCAAATA 2081
Db      239 TTTTATTTATATTTATATTAATAATTTATTAATTTTAACTTTTAAAAATTTTATATATAT 180
Oy      2082 TCTCATTAATATTTTGTGAAATCTACAAATTTATTAATTTAGTCAATATACATGACATA 2141
Db      179 TTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAA 120
Oy      2142 GAAAGTCCAAAAAATTTTGTATACAGAACTCCAAATTTTATTTTATGACACA 2201
Db      119 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60
Oy      2202 AGAATATACAGATAGAAA 2220
Db      59 AAAAAAAAAAAAAAAAAAAAA 41

RESULT 7
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: MATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENE DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match      2.0%; Score 64.2; DB 10; Length 640681;
Best Local Similarity 46.4%; Pred. No. 1.2;
Matches 327; Conservative 0; Mismatches 363; Indels 15; Gaps 3.

Oy      1631 GTGATTAATCCAAATACACTCTTGAGGTGTTAAACCTTAATCTGCTCTCCACGGTA 1690
Db      609568 GTCATCTTCAATCATACAGATTTAAATTTTATATGATTTTCATATATTTTCATAAT 609509
Oy      1691 TAAATATTAATCAGGAGGTAAAAAAGTTTGCTATTTTCGGGATCCATGAAGATA 1750
Db      609508 TATTAATAATTTTTTTTATTAACATTAACCTGATTTTATTAAGATTAATAAGAAATTTT 609449
Oy      1751 AACCTAATGACTTTATTTTGGAAATAGTAAACCTTTACATAGATTAATACCGTA 1810
Db      609448 TTAATATTAATTAATATGATTTTTTTATTTATTAATTTCTATATATATATATATATA 609389
Oy      1811 TGTATTTGTGCCAATATGACAGCCTCTACAACTGTGATAGTCAATTTTTTCGCAATA 1870
Db      609388 TAAAAATTTTTTGGATAGATATATATTTCTGAAAAAATTTTAAATATTTTATATACACA 609329
Oy      1871 TTAATTTAGGAATTCAAATGCTACTATCATATGACAGAAACAGCTGAGTATTAATTTAAT 1930

```

Query Match	2.0%	Score 62.8;	DB 9;	Length 335913;
Best Local Similarity	47.3%;	Pred. No. 1.7;		
Matches 258;	Conservative 0;	Mismatches 283;	Indels 6;	Gaps 2
QY 1923	ATTTTAAATTTAAAGACAAAAATTTTGGAAAAAGTTATATATTTCTTACCAATATATTATTTAAA	1982		
Db 280434	ATTTCGTGATTGTATTAACGCGTAATTAATTAACCTAACGAAATATATTAATTTTAAATATAA	280375		
QY 1983	TATGATGGCTAAATAGTATTTCCATATGTTCTTAAATAATTTTTTTTATATTAGTTATA	2042		
Db 280374	TTTACCCATTAACATAGATATGTATTAATTAATTTATATTCACAAATATGATTTATATATA	280315		
QY 2043	AATACATTATGACCAATATAGTTGGTGAATTCACAAATCCCATTAATATTTTGGAA	2102		
Db 280314	ATAAATATAAATATATATACTAATATCTAATTTATATAACATATATATATATTTATATCTTTTA	280255		

Query Match	2.0%	Score 62.8;	DB 9;	Length 335913;
Best Local Similarity	47.3%;	Pred. N1.7;		
Matches 258;	Conservative	0;	Mismatches 282;	Indels 6;
			Gaps	2
QY 1932	ATTTTAATTTAAAGCAAAAATTTTGGAAAAAGTATATATTTCTACACATATATTATAAA	1982		
DB 280434	ATTTCTGATATAGTTAAACAGGTAAATTAACCTACAAACGAAATATATATTTTAAATAAA	280375		
QY 1983	TATGATGCGCTAATAAAGTATTTCCATATGTTCTTAAATAATTTTTTTTATTTAGCTATA	2042		
DB 280374	TTTACCTATTACATAGATATATGATATATATTTATTTATGCAAAATGTGATTTATATATA	280315		
QY 2043	AATACATTATGAAACCAATATAGTTGTGCAATTCACAAATCCCAATATATTTTGGAA	2102		
DB 280314	ATAAATATATAATATATACTAATACATATTTATATTAACATATATATATATTTATATCTTTTA	280255		

	Query Match	Score 62.2%	DB 9	length 127197
	Best Local Similarity	43.5%	Pred. No. 1.5'	
	Matches 383	Conservative	0	Mismatches 493
			Indels 5	Gaps 2
QY 1371	ATCACTGCAAAATATCATCTCTGGTGGTCACACATPAAAAACAACAAAAAAGAA	1430		
Db 71604	ATCATTAATTCACAAATTCATTTTATATAAATATATACAAAATCTAAATTTATCTA	71663		
QY 1431	AACGATTTTTCTGGATTCATTCATGATCTPAAAAATGCATGATCTTTGGGTACAGT	1490		
Db 71664	AA--ATTATTATATTTTGTGTTATATAGATTAAAAATGATCTATATGAAATTTTAAAA	71721		
QY 1491	TTGCAAGTCTCTACACGCTGTAAACATCTGCAACTATPAAATGGCTTCTTTAATGCA	1550		
Db 71722	TATAAATATTCGATTAATTTTTTATGTGTGTCATTTTAAAAATTTGATTCACAAATGA	71781		
QY 1551	TCTTTAACATTTTTTTCTTTAGT---GGAATTTAATAAGACGAAGCTGTAACTTACA	1607		
Db 71782	TAGTATATATATATAGATATAATTTTAACGATTAGATTAAATAAAAAATATATTTATAT	71841		

Query Match	1.98;	Score 61.6;	DB 9;	Length 3410;
Best Local Similarity	45.48;	Pred. No. 0.49;		
Matches 305; Conservative	0;	Mismatches 359;	Indels 8;	Gaps 2

```

QY 1550 ATCTTACATATTTATTTAGTTGCAATTAATAAGCAGAACTTGATACAT 1609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 569 AGCTTAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 628
QY 1610 ATTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 1669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 TTTACCTGATGATGATGATGATGATGATGATGATGATGATGAT 681
QY 1670 ATCTGTTCTGCTGATGATGATGATGATGATGATGATGATGAT 1729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 682 TTTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 741
QY 1730 TTTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 1789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 742 TTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 801
QY 1790 ACTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 802 AATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 861
QY 1850 AGTCAATTTTCTGCAATATTAATTAATTAATTAATTAATTAAT 1909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 862 TTTTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 921
QY 1910 AGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 922 AATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 981
QY 1970 AATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2029
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 982 ATCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1041
QY 2030 ATATGATGATGATGATGATGATGATGATGATGATGATGAT 2088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1042 TTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1101
QY 2089 TATATGATGATGATGATGATGATGATGATGATGATGATGAT 2148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1102 TAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1161
QY 2149 CCAAAAAATTTTGTAAACAGAACTTCAAAATTTTGTAAACAGAA 2208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 TTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1221
QY 2209 ACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1222 CATTTAAGTAAA 1233

```

RESULT 12

```

US-09-938-842A-4756
; Sequence 4756, Application US/0993842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4756
; LENGTH: 1713

```

```

; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4756

```

```

Query Match 1.9%; Score 61.4; DB 9; Length 1713;
Best Local Similarity 46.6%; Pred. No. 0.41;
Matches 229; Conservative 0; Mismatches 261; Indels 1; Gaps 1;

```

```

QY 1928 AATTGATGATGATGATGATGATGATGATGATGATGATGAT 1987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 AATTGATGATGATGATGATGATGATGATGATGATGATGAT 710
QY 1988 TGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 2047
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 AATGATGATGATGATGATGATGATGATGATGATGATGATGAT 770
QY 2048 AATTGATGATGATGATGATGATGATGATGATGATGATGAT 2107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 AATTGATGATGATGATGATGATGATGATGATGATGATGAT 830
QY 2108 CAATGATGATGATGATGATGATGATGATGATGATGATGAT 2167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 GAATGATGATGATGATGATGATGATGATGATGATGATGAT 889
QY 2168 CAGAACTTCAATTTTGTAAACAGAACTTCAATTTTGTAAAC 2227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 890 TTTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 949
QY 2228 GTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 950 AATTGATGATGATGATGATGATGATGATGATGATGATGAT 1009
QY 2288 ACAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1010 AATTGATGATGATGATGATGATGATGATGATGATGATGAT 1069
QY 2348 TGGGATGATGATGATGATGATGATGATGATGATGATGATGAT 2407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1070 ATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1129
QY 2408 AATGATGATGATGATGATGATGATGATGATGATGATGAT 2418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1130 ACTTGTATTC 1140

```

RESULT 13

```

US-09-774-414-2
; Sequence 2, Application US/09774414
; Patent No. US20020102231A1
; GENERAL INFORMATION:
; APPLICANT: The Institute of Physical and Chemical Research
; TITLE OF INVENTION: Endonuclease
; FILE REFERENCE: PH-651
; CURRENT APPLICATION NUMBER: US/09/774,414
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 09/306,970
; PRIOR FILING DATE: 1999-05-07
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1428)
US-09-774-414-2

```

```

Query Match 1.9%; Score 61.2; DB 10; Length 1431;
Best Local Similarity 47.4%; Pred. No. 0.41;
Matches 215; Conservative 0; Mismatches 238; Indels 1; Gaps 1;
QY 1899 ATGAGAAAGACGATGATGATGATGATGATGATGATGATGAT 1958

```

Db	164	ATATAATAAAATTTATATATATATATGATATGATATATATATGTAATATATATATACATCATATA	223
Qy	1359	TAATTTCTAAACAATATATATTAATAATATGATGCCATATATGTAATTTCCATGTTCTTAAAA	2018
Db	224	AACACGTACACCTATGCTCTAATTAATAAAGATTAATGATTTCAAAAAATATTAATGATTTATA	283
Qy	2019	TATTTTTTTTATATTTAGTATATTAATTAATATATGACCAATATATAGTTGGGATTTCAA	2078
Db	284	AATTTATATTAATCTATTTTATATATTTTAAATAAAAATGAAATGGATATGTAATATA	343
Qy	2079	ATATCTCCATTAATATTTTTTATGGAATCTPAC - AATATTAATATTTAGCATTAATCATG	2137
Db	344	ATTAATATATTAATTAATTAATATTTCTTTAAATATATATGATATATTAATAAATATTAAGAAATA	403
Qy	2138	CATGAAGAAGTCCAAAAAAATTTGTTAAACAGAAACTCCAAATTTTTTTTTTTTTTATGG	2197
Db	404	ATTTAAATTTAAAAACCTCTAATATATGAAACATTAATTTACTATATATTTTTTTTAAATGG	463
Qy	2198	AACAGAATAATACAGATAGAAAACATAATTTGTTGGTAAGGAAGTAGTAATATACATTA	2257
Db	464	ATAAATATTTTAATAATTAATAATATATGAATAATTTAGATATTTTAATATATGATTCCTATA	523
Qy	2258	AGCAAAATTTAAAAAATATATATAGCCTATAGCGCTCAAAGTATGTTATCTAGTAGGTG	2317
Db	524	ATTATATGTTTAATATATATTAATTAATTAAGATTAATTAATTAATTAATTAACAGATATAGATT	583
Qy	2318	TAATTAATATATGATGTCGATATTCAGAAATTTGGG	2351
Db	584	TAAATTAATATGAATTTTATGATTAATTTATTCAGG	617

```

: RESULT 14
: US-09-962-832-154
: Sequence 154, Application US/09962832
: Patent No. US20020110821A1
: GENERAL INFORMATION:
: APPLICANT: Ebner, Reinhard
: TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signature
: TITLE OF INVENTION: Sels
: FILE REFERENCE: 689290-74
: CURRENT APPLICATION NUMBER: US/09/962,832
: PRIORITY FILING DATE: 2001-09-25
: PRIOR APPLICATION NUMBER: US/60/235,077
: PRIOR FILING DATE: 2000-09-25
: PRIOR APPLICATION NUMBER: US/60/235,280
: PRIOR FILING DATE: 2000-09-25
: NUMBER OF SEQ ID NOS: 259
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 154
: LENGTH: 302250
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-962-832-154

```

[illegible][illegible]

RESULT 15
 US-09-938-842A-3532/C
 Sequence 3532, Application US/09938842A
 Patent No. US20020160378A1
 GENERAL INFORMATION:
 APPLICANT: Harper, Jeff
 APPLICANT: Kreps, Joel
 APPLICANT: Wang, Xun
 APPLICANT: Zhu, Tong
 TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 TITLE OF INVENTION: SAME, AND METHODS OF USE
 FILE REFERENCE: SCRIPT300-3
 CURRENT APPLICATION NUMBER: US/09/938, 842A
 CURRENT FILING DATE: 2001-08-24
 PRIOR APPLICATION NUMBER: US 60/227, 866
 PRIOR FILING DATE: 2000-08-24
 PRIOR APPLICATION NUMBER: US 60/264, 647
 PRIOR FILING DATE: 2001-01-16
 PRIOR APPLICATION NUMBER: US 60/300, 111
 PRIOR FILING DATE: 2001-06-22
 NUMBER OF SEQ ID NOS: 5379
 SEQ ID NO 3532
 LENGTH: 2000
 TYPE: DNA
 ORGANISM: Arabidopsis thaliana
 US-09-938-842A-3532

Query Match	1.9%	Score 60.8	DB 9	Length 2000
Best Local Similarity	47.0%	Pred. No. 0.54		
Matches 186	Conservative	0	Mismatches 212	Indels 0
			Gaps	
QY 1950	AAATGTTAATTTCTCAACAATATTTAAATATGATGACCTATATGATATTTCCATG	2009		
Db 811	AAATTTCTACAGACACCAAGAGGACCAAAAAATTAATTCGTATTTATTAATCAAA	752		
QY 2010	TTCTTAATAATTTTTTTTATTTATTTAGTTATTAATACATTATGAAACCAATTAAGTTGG	2065		
Db 751	TTTAAAAACCTTATGATATATTTATATTTCTATATTAATTAATTTCTATATCTTTAG	692		
QY 2070	TGAATTCAAATATCTCCATTAATTTTTTTGAAATCTACAATTAATTAATTTAGTCA	2139		
Db 691	TATATATGATATATATATATATATATATATACGTTTTTTATCAACTTTCATTAATGAAATG	632		
QY 2130	TACAAATGCATAGAAAGTCTCCAAAAAAATTTGTTACAGAACTTCCAAATTTTTTT	2188		
Db 631	TATCAATTTATTTTAAATTTAGTAAATTAATGTTTTTAAAGGAAATTTATTTAAATTT	572		
QY 2190	TTTTATGACAAGAAATAACATATGAATGAACATTTTCTGTGGAATGAGATGATAT	2249		
Db 571	GTAATATATTTTAATTTTTTACATCTATTAATAATTAATTTTGAAGGAAAAATTTATAT	512		
QY 2250	ATACATTACCAAAATTTTAAAAAATTAATTAAGCCTATAGCGGCTCAAGATATGTTACT	2309		
Db 511	TTAATTTATTTTAAATTCATACAGAACTTAATTTTAAATGAATTAATTAACAAATTC	452		
QY 2310	AGTAGTGCTAATTAATATGATGAGGGCGCATTCAGAAATG	2349		
Db 451	AATTAATGATTTTTTATGATATTTCTATATTTGTAATTTTATG	412		

```
Search completed: March 30, 2003, 13:11:02
Job time : 4203.88 secs
```

Mon Mar 31 09:27:33 2003

us-09-502-426a-1_copy_1_3202.rnpb

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:08:52 ; Search time 4178.19 Seconds

(without alignments)
12411.569 Million cell updates/sec

Title: US-09-502-426a-1_COPY_1_3202

Perfect score: 3202
Sequence: 1 atgttggtatcatattgttgc.....gagagagagaactagctcc 3202

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estow:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_iny:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195	6.1	473	17	AL768954 Arabidopsis
2	148	4.6	148	17	BH811044 Arabidopsis
3	146.4	4.6	473	17	AL768954 Arabidopsis
4	113.2	3.5	1101	17	CNS00EVL
5	110.8	3.4	1101	17	CNS00EVL
6	109.6	3.4	1101	17	CNS0021T

C	7	98.4	3.1	836	17	CNS01100
C	8	98.2	3.1	1101	17	CNS00E07
C	9	98	3.1	807	10	AV717372
C	10	98	3.1	1092	17	CNS020K7
C	11	94	2.9	1101	17	CNS003BD
C	12	94	2.9	1201	17	CNS0167M
C	13	93.6	2.9	1101	17	CNS003BD
C	14	93.6	2.9	1101	17	CNS00E07
C	15	93	2.9	1101	17	CNS00BO1
C	16	92	2.9	661	17	CNS020Y7
C	17	91.8	2.9	886	17	BH177277
C	18	91.8	2.9	886	17	CNS070X
C	19	91.8	2.9	1101	17	CNS0039G
C	20	90.6	2.8	945	17	CNS04D0K
C	21	90.6	2.8	1187	17	B11102
C	22	90.4	2.8	836	17	CNS01100
C	23	89.6	2.8	1029	17	CNS01ZGM
C	24	89	2.8	987	17	CNS014P0
C	25	88.6	2.8	1092	17	CNS020K7
C	26	88.6	2.8	1190	17	CNS020M7
C	27	88.4	2.8	1190	17	CNS0039G
C	28	88.2	2.8	1101	17	CNS0016D
C	29	87.4	2.7	1169	17	CNS06RHO
C	30	87	2.7	1225	17	CNS0161D
C	31	86.8	2.7	1101	17	CNS0021J
C	32	86.6	2.7	1101	17	AO258984
C	33	86	2.7	1013	17	CNS06RHO
C	34	85.8	2.7	1187	17	B11102
C	35	85.6	2.7	961	17	CNS008HI
C	36	85.2	2.7	928	17	CNS00DKY
C	37	85	2.7	1101	17	CNS05AB2
C	38	84.8	2.6	1084	17	CNS072NH
C	39	84.8	2.6	1101	17	CNS003BH
C	40	84.6	2.6	1190	17	CNS020N7
C	41	84.6	2.6	804	17	B12681
C	42	84.4	2.6	1101	17	CNS006RQ
C	43	84.4	2.6	1101	17	CNS00BO1
C	44	83.8	2.6	609	17	CNS025K2
C	45	83.8	2.6	886	17	BH177277

ALIGNMENTS

RESULT 1
LOCUS AL768954
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-082A08-011867,
genomic survey sequence.
ACCESSION AL768954
VERSION AL768954.1 GI:21522073
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
Strizhov, N., Li, Y., Rosso, M., Viehoveer, P., Dekker, K., Saedler, H.
and Weisshaar, B.
A pipeline for automated high-throughput generation of ESTs
(flanking sequence tags) from Arabidopsis thaliana T-DNA
transformed lines
Unpublished
2
Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.
A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
for flanking sequence tag based reverse genetics
Unpublished
3 (bases 1 to 473)
Li, Y., Rosso, M., Strizhov, N. and Weisshaar, B.
Direct Submission
Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer

TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Weisshar B., Max-Planck-Institut fuer
Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany

COMMENT This sequence is recovered from the left border of the T-DNA. It
indicates an insertion within the locus defined by clone T3A5. The
sequences are generated at the MPI for Plant Breeding Research in
the context of the GABI-Kat project. GABI-Kat is part of the German
Plant Genomics program designated 'GABI'. Information on line
availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES
source
1.473
Location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-082A08-011867"
/note="PCR was performed on DNA from Arabidopsis thaliana
plants (T1) which were transformed with the T-DNA from
vector pAC161. The lines contain one or more T-DNA
insertions. The DNA fragment(s) resulting from the PCR
were directly sequenced to determine the genomic sequence
flanking the insertion. Sequences displaying significant
similarity to the A. thaliana nuclear genome sequence were
processed for submission. T-DNA derived sequences were
removed"

BASE COUNT 168 a 58 c 70 g 146 t 31 others
ORIGIN

Query Match 4.6%; Score 146.4; DB 17; Length 473;
Best Local Similarity 90.8%; Pred. No. 1.2e-11;
Matches 167; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 1624 CTAGATGATGATTTCCAAATACATCTT-TGGATGTTAACTTAATCTTCTTC 1682
|||||
DB 186 CTGATGATGATTTCCAAATACATCTTGTGATGTTTAACTCAATCTGTTCTT 127
|||||
QY 1683 CTAGGATTAATATTAATCATCGAGGATTAAGTTTGTCTATTTTGGCGATGCAT 1742
|||||
DB 126 CCACTGATTAATATTAATCATCGAGGATTAAGTTTGTCTATTTTGGCGATGCAT 67
|||||
QY 1743 GAAGGATTAACCTTAATGACTTAATTTTGAAGAAAGTAACTTATTCATATGATTA 1802
|||||
DB 66 GAAGGATTAACCTTAATGACTTAATTTTGAAGAAAGTAACTTATTCATATGATTA 7
|||||
QY 1803 TTAC 1806
|||||
DB 6 TTAC 3

RESULT 4
CNS00EVL/C 1101 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BAC29823 of RPL-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
genoscope.

AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Piter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPL-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1.1101
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR29823"
/clone_1lb="RPL-98"
/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 3.5%; Score 113.2; DB 17; Length 1101;
Best Local Similarity 35.8%; Pred. No. 4e-07;
Matches 232; Conservative 118; Mismatches 297; Indels 1; Gaps 1;

QY 1625 TAGATGATGATTTCCAAATACATCTTGGATGTTAACTTAATCTTCTTCCT 1684
|||||
DB 1100 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1041
|||||
QY 1685 ACGGATTAATATTAATCATCGAGGATTAAGTTTGTCTATTTTCCGATGATGA 1744
|||||
DB 1040 WATATATTAATTTTAAATATATATATATATATATATATATATATATATATAT 981
|||||
QY 1745 AGGATTAACCTTAATGATTTTGAAGAAAGTAACTTATTCATATGATTAAT 1804
|||||
DB 980 WTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 921
|||||
QY 1805 ACCGATGATTTTGGCCATTAATGACGCTGACATGATGATGATGATGATGAT 1864
|||||
DB 920 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 861
|||||
QY 1865 CAATATTAATTAATGATTAATGATGATGATGATGATGATGATGATGATGAT 1924
|||||
DB 860 TTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 801
|||||
QY 1925 TTTAATTTTAAGCAAAATTTTGAAGAAAGTATATTTCTAACAATATTTAAATA 1984
|||||
DB 800 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 741
|||||
QY 1985 TGATGCTATTAATGATTTTCCAT-TGTTCTTAAATATTTTATTTTATTTTATTT 2043
|||||
DB 740 WTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 681
|||||
QY 2044 ATACATTAAGCAAAATTAATGATGATGATGATGATGATGATGATGATGAT 2103
|||||
DB 680 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 621
|||||
QY 2104 TCTAACAATTAATTAATTTAGCAATTAACATGCAATGCAATGCAATGCAAT 2163
|||||
DB 620 AAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 561
|||||
QY 2164 TTAACGAACAACTTCCAAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2223
|||||
DB 560 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 501
|||||
QY 2224 TTTTGTGGAATGGAAGTATTAATTAATTAATTAATTAATTAATTAATTAAT 2271
|||||
DB 500 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 453

RESULT 5
CNS00EVL

[illegible]

QY	1808	GTATGTTTGTGGCCATTAATGACAGCCTCTAACAC--TGATAGTCAATTTTTCGTG	1864
Db	816	KATWAAAMAMAMATAMWATWATATAMWATATAMWAMAAAAATTTAATATATAAATWATRAANA	875
QY	1933	AAAGACAAAATTTTGAAAAATGTTAAATTTCTAACAAATATTATTAATAATGATCGCT	1992
Db	876	AAAAATAMATWTTTWTWTTTWWAMWATATATAAMAMATAMAAAAAATAAAAAATTAAMA	935
QY	1993	ATAAGTATTTCCATGCTCTAATAATTTTATTTTATTTAGTATATAAATACATTAT	2052
Db	936	WWTWTATATATTTATTAATTAATTAATWATWATWTTWAMWTTWATWTTWATATATATW	995
QY	2053	GAACCAATATATAGTGGTGAATTCGAATATTCACATTAATTTTGTGAACCTACAAT	2112
Db	996	KTAATWATATATATTTATTAAMWATATTTTANAAAAATATATATATATAMWATAMATATW	1055
QY	2113	TATTAATTTAGTCATTAACATGCAATGCAAGTCCAAAAAATA 2158	
Db	1056	AWMAATTAATTTATATATWATWATWMAAAAAAAMAMATAMWAA 1101	
RESULT 6	CNS0021J/C	1101 bp	DNA
LOCUS	CNS0021J	1101 bp	DNA
DEFINITION	Drosophila melanogaster genome survey sequence TE13 end of BAC # BAC05N11 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL061936	GI:4940214	
VERSION	AL061936.1	GI:4940214	
KEYWORDS	GSS.		
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pezomyzeta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : segeif@genoscope.cns.fr)		
COMMENT	- Web : www.genoscope.cns.fr) determination of this BAC-end and sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Airon Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://BACPAC.med.buffalo.edu/drosophila_bac.htm .		
FEATURES	location/Qualifiers		
source	1..1101		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone="BACR05N11"		
	/clone_id="RPCI-98"		
	/note="end : TE13"		
BASE COUNT	631 a 7 c 28 g 289 t	146 others	
ORIGIN			
Query Match	3.4%	Score 109.6; DB 17; Length 1101;	
Best Local Similarity	43.9%	Pred. No. 1.3e-06;	
Matches 236; Conservative 65; Mismatches 230; Indels 6; Gaps 2;			
QY	1748	ATAAACCTAATGACTTATTTTGAAGAAATTAACCCCTTTACTCATAGATTAATACC	1807
Db	1097	AAAAAMCAMWATHTTTAAACCTCMTHHTTTTMTTMTCTCTTTTTTTTTTTTTT	1038
QY	1808	GTATGTTTGTGGCCATTAATGACAGCCTCTAACAC--TGATAGTCAATTTTTCGTG	1864

Query Match	3.18;	Score 98.4;	DB 17;	Length 836;
Best Local Similarity	45.38;	Pred. No. 5.8e-05;		

collaboration with the Berkeley Drosophila genome project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutomo Osoegawa and Aaron Mammotser in Pietr de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCR-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the

OY 2237 TCGAAGTAAATACATTAAGCAAAATTTTAAAAATT 2275
 DB 113 TGAATATGAAAAAATGAAATTAATTAATTTT 75
 RESULT 10
 CNS020K7/C
 LOCUS
 DEFINITION CNS020K7 1092 bp DNA linear GSS 12-MAY-2000
 Tetraodon nigroviridis genome survey sequence T7 end of clone
 222111 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 AL175696.1 GI:7813753
 GSS; genome survey sequence.
 Tetraodon nigroviridis.
 Tetraodon nigroviridis.
 Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 1092)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
 Sautin, W., and Weissbach, U.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 2 (bases 1 to 1092)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Sautin, W., Bernot, A. and
 Weissbach, U.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 3 (bases 1 to 1092)
 Genome.
 Direct Submission
 Submitted (12-APR-2000)
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 Location/Qualifiers
 source
 1. 1092
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="222111"
 /clone_1lb="G"
 /note="Genoscope sequence ID : C0AG222CF06LPI-end : T7"
 BASE COUNT 383 a 169 c 165 g 262 t 113 others
 ORIGIN
 Query Match 3.1%; Score 98; DB 17; Length 1092;
 Best Local Similarity 43.3%; Pred. No. 5.7e-05;
 Matches 176; Conservative 58; Mismatches 170; Indels 2; Gaps 1;
 OY 1762 TTTAATTTTGAATGTAACCTTCTCATAGATTAATACCGATGTTTGTG 1821
 DB 1046 TTTTATTTTATTAATTAATTAATTTTATTTATTAATTTTATTTTATTTT 987
 OY 1822 CCATAATGACAGCCCTACAACTGTGATGCAATTTTCTGCAATATTAATTAAGA 1881
 DB 986 AAAATTAATTAATTAATTTTATTAATTTTATTTTATTAATTAATTAATTAAGA 927
 OY 1882 ATTCATGCTACTATCAATTAAGAAAGACGCTAGATTAATTAATTTAAAGCAAA 1941
 DB 926 AAAAATTTAAAMWTTAAAMAAAAAAAMWTTTATTTTATTTTATTTTAAATTTWA 867
 OY 1942 ATTTTGAATAATGTTT-ATAATTTCTAACATATTAATTAATTAATGATGCTATATGT 1999
 DB 866 ATTATTAATTAATTTTAAATTAATTTTATTTTATTAATTAATTAATTTTATTTTAA 807

OY 2000 ATTCTCATGTTCTTAAATTAATTTTATTAATTTAGTTAATAATACATTAAGACCA 2059
 DB 806 ATTAAMWTTATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 747
 OY 2060 TATTAATGCTGATCAATCAATATCCATTAATTTTGAATCAATCAATTAATTAAT 2119
 DB 746 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 687
 OY 2120 ATTACTCAATCAATCAATGATGCAATTAATTTTCAAAAAAATTTGTT 2165
 DB 686 ATTTTATTTTAAAMWTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 641
 RESULT 11
 CNS003BD
 LOCUS
 DEFINITION CNS003BD 1101 bp DNA linear GSS 03-JUN-1999
 Drosophila melanogaster genome survey sequence TERT end of BAC #
 BACH08K08 of RPI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 AL064091
 AL064091.1 GI:4941847
 GSS.
 Drosophila melanogaster.
 Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genome.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see http://www.fruitfly.org The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutyo Osoegawa and
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 p1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://becpac.med.buffalo.edu/drosophila_bac.htm.
 Location/Qualifiers
 source
 1. 1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACH08K08"
 /clone_1lb="RPI-98"
 /note="end : TERT"
 BASE COUNT 395 a 120 c 103 g 334 t 149 others
 ORIGIN
 Query Match 2.9%; Score 94; DB 17; Length 1101;
 Best Local Similarity 37.0%; Pred. No. 0.00021;
 Matches 248; Conservative 80; Mismatches 342; Indels 0; Gaps 0;
 OY 1640 CCAATATACATCTTGGATGTTAACTTATGTTTCTCCTACGATTAATTAATTA 1699
 DB 432 CGATTTCTTAATHAACACTHCCAAACCCCTTTCYCHCKMCMCMMAAAAAA 491
 OY 1700 ATCATGAGGTAAAGATTTTGTCTATTTTGGGATGATGAAGATTAACCTAATG 1759
 DB 492 AATTTTAAAAAAMAAAAACCTTAATTAAMAAAAAAMAAAAAATTTTWT 551
 OY 1760 ACTTTATTTTGAATAATTAACCTTTTACATGATTAATTAATTAATTTTGTG 1819
 DB 552 WATTTTATTAATTAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMATTTT 611


```

BASE COUNT      323 a      87 c      79 g      551 t      161 others
ORIGIN

Query Match          2.9% ; Score 94; DB 17; Length 1201;
Best Local Similarity 38.4%; Pred. No. 0.0002;
Matches 247; Conservative 87; Mismatches 303; Indels 6; Gaps 2

QY   1748 ATAAACCTAATGACTTAAATTTTGGAAAAGTGAACCCTTTAGTCATGATGATTAAATACC 1807
      |||: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB   1166 ATAMMAATATATANNNAANAANTATAAAAAAATAATMWAATATAAWMAWMAWWMTAT 1107
QY   1808 GTATGTCTTTTGTCGCATATGACAGCCCTCPCCACTGTGATGACATATTTTCGCCAA 1867
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   1106 ATTAATATMAATATATANNNAANAANAANAANAANWTTHNANAATATTTTWNNT 1047
QY   1868 ATATTAAATTAAGAATTCATGCTACTATCATAGACAAACAGCGTAGATTATACATTTT 1927
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   1046 AMAATATTTTTTTTTTTTTTTTATA - WMAAAMAANAANAATTTTAAANAATAATAT 988
QY   1928 AATTTAAAGACAAAATTTTGCAAAATGTTAATTTCTAACATATATATTAATATGCA 1987
      ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   987 TATWMAAATTTTTTAAAAATTTTTTWTWTWTTTTAAAAAANAATAWMAAATTTT 928
QY   1988 TGCCCATATGATGATTCGATGCTGCTTAAATATTTTATATATAGTATATAATAC 2047
      ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   927 TTATATWTATTAAMAWMAWTTTTTTTWTATAAAAAAAAAAAMTAAAAATWRAAAATATATA 868
QY   2048 ATTATGAACAATAATAGTGTGCTGATCAATATCTCCATTAATATTTTGAATCTA 2107
      ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   867 ATAAAAAANAANAANAANAANAANAANAANMTTWATATATAWATTTTAAAAAANAANA 808
QY   2108 CAATATTTAATATTTAGTCAATAACAATGCATAGAAAAGTCCAAAANAATTTGTTAA 2167
      ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   807 AAATTTAATTTTNNAAAATAAAAAANAANAANAANAANAATWTAATATTAATTA 748
QY   2168 CAGAACTTCCAATTTTTTTTTTTTATAGAACACGAATATACATATGAAACATATTT 2227
      ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   747 ATNCATTAACCAAAAAAAMATWHTWHTAAAAAANAANAANAANAANAANAATTT 688
QY   2228 GTTGTGAATGAGAGTAGTAATATATCATTTAAGCAATTTTAAAAA----TTATATAG 2282
      ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   687 TTTTACAAAMMAAMMMYMMMCAAAAAAMMAVMAAAMMAACAMMMTTAAAGCTTAAAAA 628
QY   2283 CCTATACGCGCTCAAAGATGTTATCTAGTAGGTGTAATTAATATGCATGCTGCAATTC 2342
      ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB   627 CAAAAAANAANAANAATAAAMAAATCTAKKNDKKWDADKNKAASMBGKKMAATW 568
QY   2343 AGAATTTGGACACACATGAAGAAACGAATTTAAATTTAACTTT 2385
DB   567 ACACGTGGCGABAMAMWTGMATGGCAAMSAWSAGRRAAADT 525

RESULT 13
LOCUS       CNS003BD/c                      1101 bp      DNA           linear    GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TEJ3 end of BAC #
            BACR08K08 of RPi-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   ALD064091
VERSION     ALD064091.1 GI:4941847
KEYWORDS
SOURCE      Drosophila melanogaster.
ORGANISM   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Epiphytoidea; Drosophilidae; Drosophila.
REFERENCE   Genoscope.
            Direct Submission
AUTHORS     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
TITLE       BP 191 91006 EVAY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL
            - web : www.genoscope.cns.fr)

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:06:10 ; Search time 3525.39 Seconds

(without alignments)
2955.362 Million cell updates/sec

Title: US-09-502-426a-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagc.....tgcataatcacaacaaga 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	358	100.0	4818	8 AF044216	AF044216 Arabidops
2	356.4	99.6	84196	8 AT73A5	AL132979 Arabidops
3	121	33.8	1691	8 AF412114	AF412114 Arabidops
4	58	16.2	5449	6 AX346543	AX346543 Sequence
5	58	16.2	7560	6 AX346125	AX346125 Sequence
6	57.2	16.0	173786	2 AC107582	AC107582 Rattus no
7	56.6	15.8	309233	2 AC098557	AC098557 Rattus no
8	56.4	15.8	5926	6 AX346554	AX346554 Sequence
9	55.6	15.5	12356	6 AX251264	AX251264 Sequence
10	55.6	15.5	169546	2 AC004157	AC004157 Plasmodiu
11	55	15.4	7351	6 AX344930	AX344930 Sequence
12	55	15.4	20486	6 AX281500	AX281500 Sequence
13	54.8	15.3	93491	2 AC116967	AC116967 Dictyoste
14	54.8	15.3	349980	6 AX344570	AX344570 Sequence
15	54.6	15.3	123280	2 AC117076	AC117076 Dictyoste
16	54.6	15.3	268147	2 AC116966	AC116966 Dictyoste
17	54.2	15.1	2814	6 AX347029	AX347029 Sequence
18	54.2	15.1	6657	3 AC114263	AC114263 Dictyoste
19	54.2	15.1	67919	2 AC123610	AC123610 Mus muscu
20	54.2	15.1	204652	2 PFMAL13P6	AL049183 Plasmodiu
21	54	15.1	7138	6 AX323841	AX323841 Sequence
22	54	15.1	187013	2 AC116920	AC116920 Dictyoste
23	53.2	14.9	6047	2 AC115582	AC115582 Dictyoste
24	53.2	14.9	14147	6 AX251501	AX251501 Sequence
25	53.2	14.9	14147	6 AX347392	AX347392 Sequence
26	53.2	14.9	14147	6 AX349113	AX349113 Sequence
27	53	14.8	7261	6 AX251434	AX251434 Sequence
28	53	14.8	133501	2 AC116936	AC116936 Dictyoste
29	53	14.8	349980	6 AX344567	AX344567 Sequence
30	52.8	14.7	5678	6 AX346040	AX346040 Sequence
31	52.8	14.7	107739	2 AC116979	AC116979 Dictyoste
32	52.8	14.7	256172	2 AC005139	AC005139 Plasmodiu
33	52.8	14.7	310779	2 AC005140	AC005140 Plasmodiu
34	52.6	14.7	6657	3 AC114263	AC114263 Dictyoste
35	52.6	14.7	24091	2 AC014610	AC014610 Dictyoste
36	52.6	14.7	106993	3 AE002751	AE002751 Drosophi1
37	52.6	14.7	349980	6 AX344564	AX344564 Sequence
38	52.4	14.6	150903	2 AC113218	AC113218 Rattus no
39	52.2	14.6	6418	6 AX251768	AX251768 Sequence
40	52.2	14.6	6418	6 AX345224	AX345224 Sequence
41	52.2	14.6	16750	6 AX251068	AX251068 Sequence
42	52.2	14.6	16750	6 AX345423	AX345423 Sequence
43	52.2	14.6	66441	3 PFMAL1P4	AL031747 Plasmodiu
44	52.2	14.6	234112	3 PFMAL1P2	AL035475 Plasmodiu
45	51.8	14.5	627	11 CNS06EMH	AL395655 T7 end of

ALIGNMENTS

RESULT 1
AF044216 4818 bp DNA linear PLN 25-JUN-2001
LOCUS Arabidopsis thaliana steroid 22-alpha-hydroxylase (DWFA) gene,
DEFINITION complete cds.
ACCESSION AF044216
VERSION AF044216.1 GI:2935341
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 4818)
Choe,S., Dillkes,B.P., Fujioka,S., Takatsuto,S., Sakurai,A. and

TITLE The DWF4 gene of Arabidopsis encodes a cytochrome P450 that mediates multiple 2alpha-hydroxylation steps in brassinosteroid biosynthesis
 JOURNAL Plant Cell 10 (2), 231-243 (1998)
 MEDLINE 98158690
 PUBMED 9490746
 REFERENCE 2 (bases 1 to 4818)
 AUTHORS Choe, S., Dikes, B.P., Azpilroz, R. and Feldmann, K.A.
 TITLE Direct Submission
 JOURNAL Submitted (22-JAN-1998) Plant Sciences, University of Arizona, Tucson, AZ 85721, USA
 FEATURES
 source Location/Qualifiers
 1. .4818
 /organism="Arabidopsis thaliana"
 /strain="Ws-2"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="8 cm from nga6"
 990. .4398
 /gene="DWF4"
 /note="dwarf4"
 /allele="dwf4-1; dwf4-2; dwf4-3; dwf4-4"
 990. 1055
 /gene="DWF4"
 join(<1133. .1353,1434. .1758,1844. .1996,2095. .2409, 2562. .2654,2746. .2824,2931. .3040,3795. .4398)
 /gene="DWF4"
 /product="steroid 22-alpha-hydroxylase"
 join(1133. .1353,1434. .1758,1844. .1996,2095. .2409, 2562. .2654,2746. .2824,2931. .3040,3795. .4040)
 /gene="DWF4"
 /function="brassinosteroid biosynthetic enzyme"
 /note="member of the cytochrome P450 superfamily; CYP90B1"
 /codon_start=1
 /product="steroid 22-alpha-hydroxylase"
 /protein_id="AAC05093.1"
 /db_xref="GI:2935342"
 /translation="MFEHHTLLPLLLPSLLSLLELLILKRRNRKRLNPPGKS GMPLEGTIGYKPYATTLGDFMOQHSKYGKTYRSLFGEPTIVSADAGLNRILO NEGRLEFECYSRISGIGLKWSMLVLDMDHRSISLNLSHARLRTILKDYERH TLFLVDSMOQNSIFSAODEAKKFTFNLMAKHMDPGEETBOLKREYVPMGCVS ALPLPGVAHKAISOAPATILKTERKMERKIDKEDEEVEKTTDEAMKSDH VRKORTDDLLGWVLSKNSLSTEQILITLSLRGHETSSVALALALFFLOACKAV EELREHEIARAKLEGESELMNDYKKMDFTCCVNETLRGIVNVEFLRKAKDY RYKGYDIPSGWKVLPVISAVHLDNSRYDQPNLFPWRMOQNNSSSGSSFTWGN NYMPFGGPPRLCAGSELAKLEMAVFIHHLVLFKFWELAEDDQFAFPVDPNGLPIR VSRIL"
 BASE COUNT 1566 a 727 c 888 g 1637 t
 ORIGIN
 Query Match 100.0%; Score 358; DB 8; Length 4818;
 Best Local Similarity 100.0%; Pred. No. 5e-51;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 AAAAAAAAAAAGATATTTTATTTCTCTCTTTTGTGATATTTTAATCA 60
 Db 4041 AAAAAAAAAAAGATATTTTATTTCTCTCTTTTGTGATATTTTAATCA 4100
 Oy 61 TTTTGGCCCAATGATATATATAAATTTGGATATAATATATGATTTGCTTT 120
 Db 4101 TTTTGGCCCAATGATATATATAAATTTGGATATAATATATGATTTGCTTT 4160
 Oy 121 TTAGTTGGGGTTTGAAGAAAGGTTTGCACCTTCGAAAGTGACGATATAGATTGG 180
 Db 4161 TTAGTTGGGGTTTGAAGAAAGGTTTGCACCTTCGAAAGTGACGATATAGATTGG 4220
 Oy 4161 TTAGTTGGGGTTTGAAGAAAGGTTTGCACCTTCGAAAGTGACGATATAGATTGG 4220
 Db 181 GAGCAGAGTTGAGCTTTGGACATTTGATTTGATTTGATTTGATTTGATTTG 240
 Db 4221 GAGCAGAGTTGAGCTTTGGACATTTGATTTGATTTGATTTGATTTGATTTG 4280
 Oy 241 ATTAAACCTTAATGGCTTCTATAGGCCCAATTTATATTGATTTAACAAGTGAC 300
 Db 4281 ATTAAACCTTAATGGCTTCTATAGGCCCAATTTATATTGATTTAACAAGTGAC 4340
 Oy 301 AACTTTACTCTGTTTTTGTATCGGAAGCAATTAACAATTTGTCAATACCAAG 358
 Db 4341 AACTTTACTCTGTTTTTGTATCGGAAGCAATTAACAATTTGTCAATACCAAG 4398
 RESULT 2
 LOCUS AT73A5
 DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone T3A5.
 ACCESSION AL132979
 VERSION AL132979.2 GI:6782244
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 84196)
 Bloecker, H., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetler, F. and Salanoubat, M.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 84196)
 AUTHORS EU Arabidopsis sequencing project.
 TITLE Direct Submission
 JOURNAL Submitted (25-JAN-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de
 Coordinator: Marcel Salanoubat and Francis Quetler, Groupement d'Interet Public, Centre National de Sequencage - GENOSCOPE, 2 rue Gaston Creteil, BP191, 91006 Evry Cedex, France;
 http://www.genoscope.cns.fr
 On Jan 27, 2000 this sequence version replaced gi:6434247.
 Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
 COMMENT
 FEATURES
 source 1. .84196
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"
 2983. .3338
 /note="355 bp LTR"
 3339. .7564
 /note="LTR-transposon"
 7565. 7922
 /note="367bp LTR"
 12339. .13197
 /gene="T3A5.10"
 join(12339. .12665,12742. .12803,12888. .13055,13125. .13197)
 /gene="T3A5.10"
 /note="similarity to cyclin-dependent kinase inhibitor (ICK1), Arabidopsis thaliana, EMBL:AT094772.1"
 (IC1), Arabidopsis thaliana.
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB62432.1"
 /db_xref="GI:6561966"
 /translation="MAVRRRRDVEENGVTYTTVKKRKKMEVDVDSRIILSPV QATNNGIVARRSACASSETSVYIVRRRSPPEEQQLIEEDSVSCSTSEKSRK IEEVDLEENGGDRETERTSWIYDILNKSEESNMDSVAVEDVSRRLRLSHETV KEALDEDFEOVAENKRLKLECSMKYNFDEKDEPLGGREYENKLP"
 12339. .12665
 /gene="T3A5.10"
 /number=1
 12666. .12741
 /gene="T3A5.10"
 /number=1
 12742. .12803
 /gene="T3A5.10"
 /number=2
 12804. .12887
 LTR
 misc-feature
 CDS
 gene
 exon
 intron
 exon
 intron

```
/gene="T3A5.10"
/number=2
12888..13055
/gene="T3A5.10"
/number=3
13056..13124
/gene="T3A5.10"
/number=3
13125..13197
/gene="T3A5.10"
/number=4
15858..16470
/gene="T3A5.20"
complement(join(15858..15901,16014..16470))
/notes="similarity to predicted proteins, Arabidopsis
thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB62433.1"
/db_xref="GI:6561967"
/translation="MARLEHNSMNGAARISFSEFVEIRSEKSNKSNINIRSSFS
MSADPAPSVTVSMTPADEITLKGKILPEKETSIVHRTIGEELTDEEGSVDCNTF
SRPITLSSSSSTIGTWRELLGLKTRHVRKTKDKVNEEVLQDHKILISGVATREC
QVADTR"
complement(15858..15901)
/number=1
complement(15902..16013)
/number=1
complement(16014..16470)
/number=2
18303..19931
/gene="T3A5.30"
complement(18303..19931)
/codon_start=1
/product="scarecrow-like 7 (SCL7)"
/protein_id="CAB62434.1"
/db_xref="GI:6561968"
/translation="MAYWCTDGNLMAIAOLIKOKOQOOSQHOQOHOEPNPNP
PSGFTLPGSGFDPQVYNDPGRPHFLEHONAASAEEEDDEMEBULINGDAS
QTNDFPIYGHDPVFPSPRLASPSYLNRKNDASAQOLPPAPASTAIVSPSPSO
HPPPPPOPDFLNOPIKAIHDVARKPERPTDLRIKESVSSEGDPIGVSYFSE
ALSKETESPSSSSSLEDPTILSKTLNACPSKFAHLANQALILEANQSNHII
AIIIDNEFTFVLPDIOLNGSSFRVDPDEVLVNEMLELKLDETATVGTALRLA
RSLNFRIVTIGEYVSLNRYEFANRVKNSLREYSAVESELEPMLDSDSKRLVRYL
FGRIMDLVRSDDNNKPGTRFGIMEKEEQRYLMEKAGEPEYKPSNVAVSQAKLLIM
NYSYSLVSESPGFI SLAMNNVPLLTYSWR"
complement(18303..19931)
/number=1
26093..28999
/gene="T3A5.40"
complement(join(26093..26338,27091..27200,27308..27386,
27478..27570,27723..28037,28136..28288,28374..28698,
28779..28999))
/notes="Contains Cytochrome P450 cysteine heme-iron ligand
signature AA455-464:Prokaryotic membrane lipoprotein lipid
attachment site AA320-330"
/codon_start=1
/product="steroid 22-alpha-hydroxylase (DWF4)"
/protein_id="CAB62435.1"
/db_xref="GI:6561969"
/translation="MFEYHTHTLLPLLPISLILFLILKRNKRRTRENLPPGKS
GMPFLGETTIGVLPKPYATLIDGPMOAHVSKYKTYRSLFGEPTIVSADAGINRILQ
NGRLEEGSPRSIGIGILGKSMVLVGDMDHMDMSISLNFISHARLTLILKDEYERH
TLFLVDSMOONSIFSNQDAKFTTNLAKHIMSDPGEETEQKKEVIFMKGVYS
APILNPGTRYHKAQSRAITLKEIRKMEERLIDIKEDOEDEEYKTEDEABMSGDH
VKQRTDDDLGWLKHSNLSTEQILDLILSLFAGHESVVALALATEFLOACQKAY
EBLREHLEIARAKKELESELMNDYKMKMPTQCVINETRLIGVAVTELRKALADV
RYKGYDIPSGMWKVLPIVSAVHLNDSRYDOPNLFNWRQOONNGSSGSGSFTSWN
NYMPGGGPRLCAGSELAKLEAVYIHLVLKFNWELAEEDNKPFAFPVDPNPGIPIR
VSRIL"
complement(26093..26338)
/number=1
```

```
intron complement(26339..27090)
/number=1
exon complement(27091..27200)
/number=2
intron complement(27201..27307)
/number=2
exon complement(27308..27386)
/number=3
intron complement(27387..27477)
/number=3
exon complement(27478..27570)
/number=4
intron complement(27571..27722)
/number=4
exon complement(27723..28037)
/number=5
intron complement(28038..28135)
/number=5
exon complement(28136..28288)
/number=6
intron complement(28289..28373)
/number=6
exon complement(28374..28698)
/number=7
intron complement(28699..28778)
/number=7
exon complement(28779..28999)
/number=8
38307..41323
/gene="T3A5.50"
complement(join(38307..38796,38877..39090,39189..39270,
40180..40242,40337..40401,40489..40564,40655..40734,
40805..40919,41225..41323))
/notes="Contains Eukaryotic putative RNA-binding region
RNP-1 signature AA179-186"
/codon_start=1
/product="U1 snRNP 70K protein"
/protein_id="CAB62436.1"
/db_xref="GI:6561970"
/translation="MDSQDPEPLRNPNMAVQARAKVONRANVLTOLKMGOSHPTG
NLKLEPEPPPEYKPEPKRCPYTGACQFVSNABEGDEYAPKPYRELDSQK
EPIHKLRIKGVKAEADLKRPDNNPAATGDPYKTLFSLRYTESSEKIKKEPES
VPIKRVHIVTQNLNPKPGYAFLEYMTTRMKAAKAYKADQKIDGRVLDVBERGT
VNMNPRIVTGGIGLGSVYSGGEIYGEQPOGKTSQSESPREERKSKKEKEE
RSRELSDHPPRRSDRREDKHNDRQGGDRDRDRDRDRDRDRDRDRDRDRDR
GSDRTRSDHDRSKKEDYEGEYEHGGGRSDRDAEYRGEPEETRGYYEDQGD
TDRYSHRTDKMEEDFRYEREKRSKESREYVR"
complement(38307..38796)
/number=1
intron complement(38797..38876)
/number=1
exon complement(38877..39090)
/number=2
intron complement(39091..39188)
/number=2
exon complement(39189..39270)
/number=3
intron complement(39271..40179)
/number=3
exon complement(40180..40242)
/number=4
intron complement(40243..40336)
/number=4
```

Query Match 99.6%: Score 356.4; DB 8; Length 84196;
Best Local Similarity 99.7%: Pred. No. 4,7e-51;
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY 1 AAAAAAAAAAGAGAGAGATTTTATCTCTCTTTTGTGATTTTAAATCA 60
db 26092 AAAAAAAAAAGAGAGATTTTATCTCTCTTTTGTGATTTTAAATCA 26033
OY 61 TTTTGTGCCCCAATGATATATAAATAATTGGATTAATATATTTGATATTCGTTT 120
```

this work. Shinozaki, K. (RIKEN GSC) and Ecker, J. R. (SSP/Salk) contributed equally to this work as PIs.

Location/Qualifiers

1. 1691

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

/chromosome="3"

/clone="RAF106-76-F19(R13551)"

/note="ecotype: Columbia"

1. 28

29. 1570

/note="steroid 22-alpha-hydroxylase (DWF4)"

/codon_start=1

/product="AT3g50660/T3A5_40"

/protein_id="AAL06567.1"

/db_xref="GI:15724348"

/translation="MPEPEHHHTLPLILPSLSLLPLILKRRNRTRENLPKPK
GMPLEGTIGLKPYTTATLDDFMOOHSGKGIYRSNLEPPTIVSADAGLNRTIL
NEGRLECSYIPRSIGGILGKWSMLVLVDGMHDKMSISLNF,SHARKYTLILKDYER
TLFLVDSWQOHSIFSADDEAKKFTFLMAKIMNSDPEEETEQIKETVFMKGVV
APLNLPGTAYHKAALOSRATILKFTERKKEERKLDIKEDQEEEEVKTDEAMSKSDI
VRKRTDDDLIGVNLKHSNLSTEGEILDLISLFLFAGHETSSVALATFPLQACPKAA
ELREBEHEIIRAKKEGSELNMDVKKMPTCCVINEETRLGCVAFPLRAKQKDA
RYKQKIDPSGKXVLPVISAVIDLNSRDQPLPFPNRRQOQNNNGSSSGSSTWG
NYMFFGGGPRCLCASELAKLEMAVFIHVLKFWNELAEDKPPAFPFVDPPNGPLPI
VSRLL"

3' UTR 1571. 1691

BASE COUNT 524 a 286 c 363 g 518 t

ORIGIN

Query Match 33.8%; Score 121; DB 8; Length 1691;
Best Local Similarity 100.0%; Pred. No. 2,7e-11;
Matches 121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGATGAAGTATTTTATTCCTCTCTTTTGGATATTTTAATCA 60
Db 1571 AAAAAAAAAAGATGAAGTATTTTATTCCTCTCTTTTGGATATTTTAATCA 1630
QY 61 TTTTGTGGCCCATGATATATATAAAATTTGGATAATATATATTTGGATATTCGTTT 120
Db 1631 TTTTGTGGCCCATGATATATATAAAATTTGGATAATATATATTTGGATATTCGTTT 1690

OY 121 T 121
Db 1691 T 1691

RESULT 4
AX346543 5449 bp DNA Linear PAT 01-FEB-2002
LOCUS AX346543
DEFINITION Sequence 1614 from Patent WO0200928.
ACCESSION AX346543
VERSION AX346543.1 GI:18494429
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1614 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES
SOURCE
1. 5449 Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)".

BASE COUNT 1146 a 90 c 1299 g 2914 t

ORIGIN

Query Match 16.2%; Score 58; DB 6; Length 5449;
Best Local Similarity 65.4%; Pred. No. 0.68;
Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY		4	AAAAAAGATGAAGAATTCTCCTGTTTGTGGATAATTAACCATTT	63
Dd		5256	AATAAAAAAAAAAAAGGATTTTTTGAAATTTTTTTTTTAATTAAGTT	5315
OY		64	TTCCTGCCCAFGATATATATAAAAATTGGATATATATGGAATTCGTTTTA	123
Dd		5316	TGTTTAAAGTTTATTAATATATATTTTTATATATTTATATCGAATTTTTTAATT	5375
OY		124	GTTCCGGTTT	133
Dd		5376	TTTTGGGTTT	5385

KEYWORDS	HTG: HTGS_PHASE1.
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 173786)
AUTHORS	Muzny,D.M., Adams,C., Adio-Onuoha,B., Ali-osman,F.R., Allen,C.

RESULT 5					
AX346125	AX346125	7560 bp	DNA	Linear	PAT 01-FEB-2002
LOCUS	AX346125				
DEFINITION	Sequence	1196 from Patent WO0200928.			
ACCESSION	AX346125				
VERSION	AX346125.1	GI:18494011			
KEYWORDS					

SOURCE	synthetic construct.
ORGANISM	synthetic construct
REFERENCE	artificial sequences.
AUTHORS	1
TITLE	Olek, A., Piepenbrock, C. and Berlin, K.
JOURNAL	Diagnosis of diseases associated with the immune system
FEATURES	Patent: WO 0200928-A 1196 03-JAN-2002;
source	Epigenomics AG (DE)
	location/Qualifiers
	1..7560

BASE COUNT	ORIGIN
2154 a	47 c 1534 g 3805 t

Query Match	16.2%;	Score 58;	DB 6;	Length 7560;
Best Local Similarity	51.6%;	Pred. No. 0.63;		
Matches 159;	Conservative	0;	Mismatches 145.	Indels 4.

conservative 0; Mismatches 145; Indels 4; Gaps 1;

2 AAAAAAAAAAGATGAAAGTATTTTATTCCTCTTTTGTGAATTTTAAATCAT 61

62 TTTTTCCTGGGCGGAGATGACGTTTATTGTTA 278

[illegible]

122 TAGTTCGGTTTGAGAAAAGGGTTTCGACCTTTTCGAAACGCGACGACCTTATTTATGTCGTATATTT 284

Db 2848 TGTTCGATTTATATAATTATTTTAGCATTTTCTATTTTAAAAATTAATTTAT 290

182 AGCTAGCTTGAGCTCTTTGGACATTTCATTCGATCT----TGTTCATTATTAGCTCGAC 237

Db 2908 TATTATTAATTAAATTTGATAAATAGTAATTAATGTGATGTATCTATTTTAAATTT 296

238 ACTATTAAACCTTAATGGGCTTCTATTAAGGCCAATTATATTACGATTATACAAGT 297

2968 ATATTTGAATTAAATAATTTTATTTTAAAGAAATATTATGAAAATATATAAAGA 302

3028	GACAACTT	305
3029	TTT	
3030	TTT	
3031	TTT	
3032	TTT	
3033	TTT	
3034	TTT	
3035	TTT	
3036	TTT	
3037	TTT	
3038	TTT	
3039	TTT	
3040	TTT	
3041	TTT	
3042	TTT	
3043	TTT	
3044	TTT	
3045	TTT	
3046	TTT	
3047	TTT	
3048	TTT	
3049	TTT	
3050	TTT	
3051	TTT	
3052	TTT	
3053	TTT	
3054	TTT	
3055	TTT	
3056	TTT	
3057	TTT	
3058	TTT	
3059	TTT	
3060	TTT	
3061	TTT	
3062	TTT	
3063	TTT	
3064	TTT	
3065	TTT	
3066	TTT	
3067	TTT	
3068	TTT	
3069	TTT	
3070	TTT	
3071	TTT	
3072	TTT	
3073	TTT	
3074	TTT	
3075	TTT	
3076	TTT	
3077	TTT	
3078	TTT	
3079	TTT	
3080	TTT	
3081	TTT	
3082	TTT	
3083	TTT	
3084	TTT	
3085	TTT	
3086	TTT	
3087	TTT	
3088	TTT	
3089	TTT	
3090	TTT	
3091	TTT	
3092	TTT	
3093	TTT	
3094	TTT	
3095	TTT	
3096	TTT	
3097	TTT	
3098	TTT	
3099	TTT	
3100	TTT	

JUN 06 1978

MAY 30 1978

RESULT 6
NC107582

LOCUS	173786 bp	DNA	linear	HTG 13-JUL
AC107582				
Rattus norvegicus clone CH230-29R17. *** SEQUENCING IN PROGRESS				

***, 92 unordered pieces.
AC107582
ACCESSION
UNION

ERKSTON
AC10/582.3 GI:21736513

```

Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GMPF
Center clone name: CH230-29B17
-----
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 1000 of reads

```

TITLE Direct Submission
JOURNAL Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT- On Jul 12, 2002 this sequence version replaced gi.18846355.

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>

JOURNAL REFERENCE AUTHORS TITLE JOURNAL	Direct Submission Unpublished 2 (bases 1 to 173786) Worley, K.C. Direct Submission Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 173786) Worley, K.C.
REFERENCE AUTHORS	

SCHEERER, S., SCOTT, G., SHEN, H., SHOOSHARI, N., SLOSSON, I.,
SODERGREN, E., SONAIKE, T., SPIRKS, A., STANLEY, H., STONE, H.,
SUTTON, A., SWALEK, T., TAYLOR, C., TAYLOR, P., TAMERISA, A., TAMERISA, K., TANG, H.,
TANSEY, J., TAYLOR, C., TAYLOR, T., TELFORD, B., THOMAS, N., THOMAS, S.,
USMANI, K., VASQUEZ, L., VERA, V., VILLALON, D., VINSON, R., WANG, O.,
WANG, S., WARD-MOORE, L., WARREN, R., WASHINGTON, C., WATLING, S.,
WILLIAMS, G., WILLIAMSON, A., WLECZYK, R., WOODEN, S., WOILEY, K.,
WU, C., WU, Y., WU, J. F., ZHOU, J., ZORRILLA, S., NELSON, D.,
WEINSTECK, G., AND GIBBS, R.

Mozodo, R. J., Lu, X., Lucier, O., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., McWhiney, E., McLeod, M. P., Meador, M., Mel, G., Metzker, M., Miner, G., Miller, Z., Mitchell, J. T., Mohabbat, K., Morgan, A., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwento, S., Oghu, M., Okunnu, G., Oreguena, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Plickens, R., Primus, E., Pu, L. L., Quiles, M., Remy, Y., Rivers, M., Rojas, A., Rotundo, K. I., Rolfe, M., Ruiz, S.,

Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Huliyil, S., Hume, J., Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., Kiny, L., Koryah, J., Kovar, C., Kratochvíl, Z., Kureishi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Michta, A., Mandy, N., Mead, B., Lewis, L. C., Lewis, L.,

Bouck, J., Bowle, S., Brieva, M., Brown, E., Brown, M., Bonalan, D.,
Buhay, C., Burch, P., Burkett, C., Burnett, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, U., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Dalla, A.L., Davis, C., Davy-Carroll, L., Dedering, D.A.,
Delaney, K.R., Deigado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Doutswalter, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Eamintahl, C., Edgar, D., Edwards, C.C., Elhajj, C., Escotto, J.,

SOURCE	ORGANISM
Norway rat, <i>Rattus norvegicus</i>	
Rattus norvegicus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
Rattus	
1 (pases 1 to 173786)	
Muzny, D. M., Adams, C., Adio-Oduola, B., Ali-Usman, F. R., Allen, C.,	
Alsbrooks, S. L., Amaral-Jorge, H. C., Are, J. R., Ayalew, M., Banks, T.,	
Barberis, J., Benton, J., Blom, K., Blom, K., Blom, K., Blom, K., Blom, K.,	

Y-6111-6468.rge

Page 5

KEYWORDS HTG: HTGS PHASE1

Assembly program: Phrap, version 0.990329
Consensus quality: 85701 bases at least Q40
Consensus quality: 94901 bases at least Q30
Consensus quality: 103067 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 92 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1044: contig of 1044 bp in length
1045 1144: gap of unknown length
1145 2264: contig of 1120 bp in length
2265 2364: gap of unknown length
2365 3650: contig of 1286 bp in length
3651 3750: gap of unknown length
3751 4927: contig of 1177 bp in length
4928 5027: gap of unknown length
5028 6190: contig of 1163 bp in length
6191 6290: gap of unknown length
6291 7306: contig of 1016 bp in length
7307 7406: gap of unknown length
7407 8489: contig of 1083 bp in length
8490 8590: gap of unknown length
8591 9778: contig of 1189 bp in length
9779 9878: gap of unknown length
9879 11183: contig of 1305 bp in length
11184 11283: gap of unknown length
11284 12388: contig of 1105 bp in length
12389 12488: gap of unknown length
12489 13543: contig of 1055 bp in length
13544 13643: gap of unknown length
13644 14803: contig of 1160 bp in length
14804 14903: gap of unknown length
14904 16009: contig of 1106 bp in length
16010 17310: contig of 1201 bp in length
17311 17410: gap of unknown length
17411 18416: contig of 1006 bp in length
18417 18516: gap of unknown length
18517 19745: contig of 1229 bp in length
19746 19845: gap of unknown length
19846 20945: contig of 1100 bp in length
20946 21045: gap of unknown length
21046 22132: contig of 1087 bp in length
22133 22232: gap of unknown length
22233 23496: contig of 1264 bp in length
23497 23596: gap of unknown length
23597 24728: contig of 1132 bp in length
24729 24828: gap of unknown length
24829 25982: contig of 1154 bp in length
25983 26082: gap of unknown length
26083 27212: contig of 1130 bp in length
27213 28336: contig of 1224 bp in length
28337 28636: gap of unknown length
28637 29865: contig of 1229 bp in length
29866 29965: gap of unknown length
29966 31115: contig of 1150 bp in length
31116 32311: contig of 1096 bp in length
32312 33411: gap of unknown length
33412 33460: contig of 1049 bp in length
33461 33560: gap of unknown length
33561 34759: contig of 1199 bp in length
34760 34859: gap of unknown length
34860 35937: contig of 1078 bp in length
35938 36037: gap of unknown length

36038 37158: contig of 1121 bp in length
37159 37258: gap of unknown length
37259 38454: contig of 1196 bp in length
38455 38554: gap of unknown length
38555 39786: contig of 1232 bp in length
39787 39886: gap of unknown length
39887 40915: contig of 1029 bp in length
40916 41015: gap of unknown length
41017 42118: contig of 1103 bp in length
42119 43243: contig of 1025 bp in length
43244 43343: gap of unknown length
43344 44733: contig of 1390 bp in length
44734 44834: gap of unknown length
44835 46847: gap of unknown length
46848 48885: contig of 2038 bp in length
48886 48985: gap of unknown length
48986 50103: contig of 1118 bp in length
50104 50203: gap of unknown length
50204 51246: contig of 1043 bp in length
51247 51346: gap of unknown length
51347 52475: contig of 1129 bp in length
52476 52575: gap of unknown length
52576 54246: contig of 1671 bp in length
54247 54346: gap of unknown length
54347 55700: contig of 1354 bp in length
55701 55800: gap of unknown length
55801 57340: contig of 1540 bp in length
57341 57440: gap of unknown length
57441 58866: contig of 1426 bp in length
58867 58966: gap of unknown length
58967 60262: contig of 1296 bp in length
60263 60362: gap of unknown length
60363 62046: contig of 1684 bp in length
62047 62146: gap of unknown length
62147 63851: contig of 1705 bp in length
63852 63951: gap of unknown length
63952 65247: contig of 1296 bp in length
65248 65347: gap of unknown length
65348 66839: contig of 1492 bp in length
66840 66939: gap of unknown length
66940 68082: contig of 1143 bp in length
68083 68182: gap of unknown length
68183 69518: contig of 1336 bp in length
69519 69618: gap of unknown length
69619 71049: contig of 1431 bp in length
71050 71149: gap of unknown length
71150 72450: contig of 1301 bp in length

Query Match
Best Local Similarity 16.0%; Score 57.2; DB 2; Length 173786;
Matches 89; Conservative 62.2%; Pred. No. 0.4;
Mismatches 54; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAATGAAATATTTATCTCTCTTTTGGATATTTAAATCA 60
DB 27835 AAAAAAAAAAATATATTTTGTGATAATTTGATAATTTGTTT 27894
QY 61 TTTTTCCTCCATGATATATATTTGATAATTTGATAATTTGTTT 120
DB 27895 ATTTTTCCTCCATGATATATATTTGATAATTTGATAATTTGTTT 27954
QY 121 TTAGTTCGGGTTGAGAAAAGG 143
DB 27955 TTAATTAANGGGGGGAAAAGG 27977

RESULT 7
LOCUS AC098557 309233 bp DNA linear HTG 12-JUL-2002
DEFINITION Rattus norvegicus clone CH330-81P10, *** SEQUENCING IN PROGRESS
ACCESSION AC098557

VERSION AC098557.6 GI:21728854
KEYWORDS HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 309233)
AUTHORS Mueny,D.M., Adams,C., Adio-Oduola,B., Ali-Adnan,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D., Bonck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabis,J., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,C.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Marindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Melker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogulu,M., Okunolu,N., Oragunde,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Qulles,M., Ren,Y., Rivers,M., Rojas,A., Rojokokan,I., Rolfe,M., Ruiz,S., Savery,G., Schermer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Tellard,B., Thomas,N., Thomas,S., Umanal,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,O., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 309233)
AUTHORS Worley,K.C.

JOURNAL Direct Submission
TITLE Submitted (24-OCT-2001) Human Genome Sequencing Center, Department
JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 309233)
AUTHORS Worley,K.C.

JOURNAL Direct Submission
TITLE Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
JOURNAL of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 11, 2002 this sequence version replaced gi:20666030.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GHV
Center clone name: CH230-81P10
Summary Statistics
Sequencing vector: Plasmid.

Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 233229 bases at least Q40
Consensus quality: 233927 bases at least Q30
Consensus quality: 234451 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 61 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1	1070	1069:	contig of 1069 bp in length
*	1170	1169:	gap of unknown length
*	1170	2523:	contig of 1354 bp in length
*	2524	2623:	gap of unknown length
*	2624	3832:	contig of 1209 bp in length
*	3833	3932:	gap of unknown length
*	3933	4955:	contig of 1023 bp in length
*	4956	5055:	gap of unknown length
*	5056	6248:	contig of 1193 bp in length
*	6249	6348:	gap of unknown length
*	6349	7523:	contig of 1175 bp in length
*	7524	7623:	gap of unknown length
*	7624	8765:	contig of 1142 bp in length
*	8766	8865:	gap of unknown length
*	8866	10089:	contig of 1224 bp in length
*	10090	10188:	gap of unknown length
*	10190	11339:	contig of 1150 bp in length
*	11340	11439:	gap of unknown length
*	11440	12630:	contig of 1191 bp in length
*	12631	12730:	gap of unknown length
*	12731	13871:	contig of 1141 bp in length
*	13872	13971:	gap of unknown length
*	13972	14978:	contig of 1006 bp in length
*	14978	15077:	gap of unknown length
*	15078	16258:	contig of 1181 bp in length
*	16259	16358:	gap of unknown length
*	16359	17421:	contig of 1063 bp in length
*	17422	17521:	gap of unknown length
*	17522	18539:	contig of 1018 bp in length
*	18540	18639:	gap of unknown length
*	18640	19856:	contig of 1217 bp in length
*	19857	19956:	gap of unknown length
*	19957	20981:	contig of 1025 bp in length
*	20982	21081:	gap of unknown length
*	21082	22354:	contig of 1273 bp in length
*	22355	22454:	gap of unknown length
*	22455	23840:	contig of 1386 bp in length
*	23841	23940:	gap of unknown length
*	23941	25442:	contig of 1502 bp in length
*	25443	25542:	gap of unknown length
*	25543	26794:	contig of 1252 bp in length
*	26795	26894:	gap of unknown length
*	26895	28026:	contig of 1132 bp in length
*	28027	28126:	gap of unknown length
*	28127	29239:	contig of 1113 bp in length
*	29240	29339:	gap of unknown length
*	29340	30567:	contig of 1228 bp in length
*	30568	30667:	gap of unknown length
*	30668	31810:	contig of 1143 bp in length
*	31811	31910:	gap of unknown length
*	31911	33252:	contig of 1342 bp in length
*	33253	33352:	gap of unknown length
*	33353	34343:	contig of 1191 bp in length
*	34344	34643:	gap of unknown length
*	34644	35643:	contig of 1000 bp in length
*	35644	35743:	gap of unknown length
*	35744	36743:	contig of 1000 bp in length

Matches 118; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 11 AGATGAAGATTTTATCTCTTTTGTATTAATTAATCAATTTTTCG 70
 DB 9051 AGATTAATGATGATTAATTTTGTAGTTTGAAGATTTTAAATGTTTTCAGT 9110
 QY 71 CCAATGATATATAAATTTGATTAATTAATTTGATTAATGTTTTCAGTGGG 130
 DB 9111 GGTGATTTATTAATTTTATTAATTAATTAAGGTTTGTATTTTATTAATG 9170
 QY 131 TTTGAGAAAAGGTTTCGACTTCGAAAGTGAGATTAATGATGAGCTAGCTG 190
 DB 9171 TATTAGATTTGTTAGTTTGTGTTTGTATTAATGTTTAAATGTTGGAATG 9230
 QY 191 GAGCTTTGACATTTGATGATGATGATGATGATGATGATGATGATGATG 232
 DB 9231 ATTTTATGTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 9272

RESULT 10
 AC004157
 LOCUS AC004157 169546 bp DNA linear HTG 12-AUG-2000
 DEFINITION Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
 PROGRESS ***, 2 unordered pieces.
 AC004157
 AC004157.8 GI:9797712
 HTG: HTGS, PHASE1.
 KEYWORDS Plasmodium falciparum.
 SOURCE Plasmodium falciparum.
 ORGANISM Plasmodium falciparum.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 169546)
 Hyman, R.W., Funu, E.L., Qin, F., Rowley, D., Mao, J., Tamaki, T.,
 Kurdi, O.B., Conway, A.B. and Davis, R.W.
 Unpublished
 2 (bases 1 to 169546)
 Plasmodium falciparum 3D7 chromosome 12
 TITLE JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL
 COMMENT

On Aug 12, 2000 this sequence version replaced g1:8810447.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 23466: contig of 23466 bp in length
 * 23467 23666: gap of unknown length
 * 23667 169546: contig of 145880 bp in length.
 FEATURES
 source
 1. 169546
 /organism="Plasmodium falciparum"
 /db_xref="taxon:5833"
 /chromosome="12"
 /clone="PFYAC293"
 /clone="3D7"

BASE COUNT 68871 a 15381 c 15705 g 68389 t 200 others

Query Match 15.5%; Score 55.6; DB 2; Length 169546;
 Best Local Similarity 51.6%; Pred. No. 0.74;
 Matches 127; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAGATGAAATTTTATCTCTTTTGTATTAATTAATCA 60
 DB 136495 AAAAAAAAAAGATGAAATTTTATCTCTTTTGTATTAATTAATCA 60
 QY 61 TTTTGTGCCCCAATATATATAAATTTGATTAATTAATTAATTTGATTTGTTT 120
 DB 136554 TTTTGTGCCCCAATATATATAAATTTGATTAATTAATTAATTTGATTTGTTT 120

DB 136555 TTTTGTGCCCCAATATATATAAATTTGATTAATTAATTTGATTTGTTTGA 136614
 QY 121 TTAGTCGGGTTTGAGAAAAGGTTTCGACTTCGAAAGTGAGCATGATTAATG 180
 DB 136615 TGAATATGAAATTAATAAACCCTATGCTATCAATTAATCAAAATATCTCCCTTGG 136674
 QY 181 GAGCTAGGTTGATGCTTTTGAGCATTTGATTAATGATGATTTGATTAATGATG 240
 DB 136675 GTAAATGAAAAAAT 136734
 QY 241 ATTTAA 246
 DB 136735 TATATA 136740

RESULT 11
 AX344930
 LOCUS AX344930 7351 bp DNA linear PAT 01-FEB-2002
 DEFINITION Sequence 1 from Patent WO0200928.
 ACCESSION AX344930
 VERSION AX344930.1 GI:18492816
 KEYWORDS synthetic construct.
 SOURCE synthetic construct.
 ORGANISM artificial sequences.
 REFERENCE 1
 Olek, A., Piepenbrock, C. and Berlin, K.
 Diagnosis of diseases associated with the immune system
 Patent: WO 0200928-A 1 03-JAN-2002;
 Epigenomics AG (DE)
 FEATURES
 source
 1. 7351
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2230 a 42 c 1230 g 3849 t

Query Match 15.4%; Score 55; DB 6; Length 7351;
 Best Local Similarity 48.0%; Pred. No. 2;
 Matches 157; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 20 TATTTTATCTCTCTTTTGTATTAATTAATCAATTTTGGCCAAATGATA 79
 DB 3765 TTTCGATTTGCTTATGCTGATTAATTAATTAATTTTATTTTGTTCGAGATA 3824
 QY 80 TATTAATAATTTGATTAATAATTAATTTGATTAATTTGATTTGATTTGATTTGATTTG 139
 DB 3825 ATTTATTTATTTATTTTAAAGTTAATTTGATTTTATTTTAAATTTTAA 3884
 QY 140 AGGGTTTCACCTTCGAAAGTGAGCATGATTAATTAATTAATTTGATTTGATTTGATTTG 199
 DB 3885 AGGTTTAAAGTTTGGGTTATGCTTTTATTAATTTGATTTGATTTGATTTGATTTGATTTG 3944
 QY 200 GACATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 259
 DB 3945 AATAATATGTTGTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 4004
 QY 260 TTCTATTAAGGCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 319
 DB 4005 TTTATATAGTAGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4064
 QY 320 ATCCAGCAATATACAAATTTGCAAT 346
 DB 4065 TTTAATAGCAATATTAATTTTAACT 4091

RESULT 12
 AX281500
 LOCUS AX281500 20486 bp DNA linear PAT 02-NOV-2001
 DEFINITION Sequence 164 from Patent WO0117376.
 ACCESSION AX281500
 VERSION AX281500.1 GI:16608755

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:05:35 ; Search time 71.5078 Seconds
(without alignments)
11274.505 Million cell updates/sec

Title: US-09-502-426a-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagt.....tgcacaaacacacacaa 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_101002:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
25: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	358	100.0	6888	21	AAA59599
2	58	16.2	5449	24	ABU33661
3	58	16.2	7560	24	ABU33223
4	56.4	15.8	5926	24	ABU33652
5	55.6	15.5	12356	22	AA546510
6	55	15.4	7351	24	ABU32028
7	55	15.4	20486	24	ABU34611
8	54.2	15.1	2814	24	ABU34127
9	54	15.1	7138	24	ABK28455

10	53.2	14.9	14147	22	AA546743	Tumour suppressor
11	53.2	14.9	14147	22	ABK33955	Human DNA for stag
12	53	14.8	7261	22	AA546670	Tumour suppressor
13	52.8	14.7	5678	24	ABU33138	Human immune syste
14	52.6	14.7	21231	23	ABU19002	Drosophila melanog
15	52.2	14.6	6418	24	ABU32322	Human immune syste
16	52.2	14.6	6418	24	AA561073	Human immune syste
17	52.2	14.6	16750	22	AA546314	Human gene regulat
18	52.2	14.6	16750	22	ABU32972	Tumour suppressor
19	51.8	14.5	6145	24	ABU32972	Human immune syste
20	51.6	14.4	17131	24	ABU32653	Human immune syste
21	51.6	14.4	13584	24	ABU32653	Human immune syste
22	50.8	14.2	5311	24	ABU33019	Human immune syste
23	50.8	14.2	8087	24	ABU32742	Human immune syste
24	50.8	14.2	13606	22	AA545457	Human immune syste
25	50.8	14.2	13606	22	AA545457	Chemically pretrea
26	50.8	14.2	13606	24	ABU33810	Tumour suppressor
27	50.8	14.2	13606	24	ABK28313	Human immune syste
28	50.6	14.1	5649	22	AA546384	DNA transcription
29	50.6	14.1	5649	22	ABK40008	Tumour suppressor
30	50.6	14.1	5649	24	ABU32849	Human chemically p
31	50.4	14.1	5163	24	ABU32848	Human immune syste
32	50.2	14.0	14615	22	AA546704	Human immune syste
33	49.8	13.9	6134	22	AA546553	Tumour suppressor
34	49.8	13.9	6739	22	ABU32738	Human immune syste
35	49.8	13.9	8866	22	AA545453	Chemically pretrea
36	49.8	13.9	8866	24	ABK28280	DNA transcription
37	49.8	13.9	15649	24	ABU70544	Chemically treated
38	49.8	13.9	113515	24	ABU34175	Human immune syste
39	49.6	13.9	9347	24	ABU49336	Human polynucleoti
40	49.6	13.9	11422	24	ABK39936	Human chemically p
41	49.6	13.9	11422	24	ABU32218	Human immune syste
42	49.6	13.9	12639	24	ABN60106	Human chemically m
43	49.6	13.9	13427	24	ABU33926	Human immune syste
44	49.4	13.8	6609	24	ABU70526	Chemically treated
45	49.4	13.8	6609	24	AA561123	Human gene regulat

ALIGNMENTS

RESULT 1	AAA59599	standard; DNA; 6888 BP.
ID	AAA59599	
XX	AAA59599;	
AC	14-NOV-2000	(first entry)
XX		
DT		
XX		
DE		DNA encoding a cytochrome P450 enzyme designated DMF4.
XX		
KW		DMF4; cytochrome P450 enzyme; brassinosteroid; 22alpha-hydroxylation;
KW		plant phenotype; cell elongation; ss.
XX		
OS		Arabidopsis sp.
XX		
EH		
FT	Key	Location/Qualifiers
FT	promoter	1..3203
FT	TATA_signal	/*tag= a
FT		3060..3125
FT	CDS	/*tag= b
FT		3203..6110
FT		/*tag= c
FT		/product= "DMF4"
FT		/note= "contains introns"
FT	exon	3203..3423
FT		/*tag= d
FT	intron	3424..3503
FT		/*tag= e
FT	exon	3504..3828
FT		/*tag= f
FT	intron	3829..3913
FT		/*tag= g

FT exon 3914..4066
 FT /*tag= h
 FT Intron 4067..4164
 FT /*tag= 1
 FT exon 4165..4479
 FT /*tag= 1
 FT Intron 4480..4631
 FT /*tag= k
 FT exon 4632..4724
 FT /*tag= 1
 FT Intron 4725..4815
 FT /*tag= m
 FT exon 4816..4894
 FT /*tag= n
 FT Intron 4895..5000
 FT /*tag= o
 FT exon 5001..5110
 FT /*tag= p
 FT Intron 5111..5864
 FT /*tag= q
 FT exon 5865..6110
 FT /*tag= r
 FT 3'UTR 6011..6468
 FT /*tag= s
 PN WO200047715-A2.
 XX 17-AUG-2000.
 XX 11-FEB-2000; 2000MO-US03820.
 XX 11-FEB-1999; 99US-0119657.
 XX 11-FEB-1999; 99US-0119658.
 XX (ARIZ-) ARIZONA BOARD OF REGENTS.
 XX Azpiroz R, Choe S, Feldmann KA;
 PI WPI: 2000-549142/50.
 DR P-PSDB; AAB07921.
 XX
 PT New isolated dwf4 polynucleotide useful for altering the phenotype of
 PT plants, for diagnostic assays and in the production of antibodies -
 XX Claim 3; Flg 10A-G; 113pp; English.
 PS
 CC The present sequence encodes a DWF4 polypeptide. The polypeptide is a
 CC cytochrome P450 enzyme that mediates multiple steps in synthesis of
 CC brassinosteroids. Specifically, it mediates multiple
 CC 22alpha-hydroxylation steps in brassinosteroid biosynthesis. The DWF4
 CC polynucleotide is used for altering the phenotype of a plant. DWF4
 CC plants display a dramatic reduction in the length of different organs,
 CC and this size reduction is attributable to a defect in cell elongation.
 CC The DWF4 polynucleotides and polypeptides can be used in diagnostic
 CC assays and to generate antibodies, which can be used to produce
 CC immunogenic compositions.
 XX
 SO Sequence 6888 BP; 2294 A; 1010 C; 1193 G; 2391 T; 0 other;
 Query Match 100.0%; Score 358; DB 21; Length 6888;
 Best Local Similarity 100.0%; Pred. No. 3.2e-55;
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AAAAAAAAAAGATGAAAGTATTTATCTCTCTTTTGGATTAATTTAAATCA 60
 DB 6111 AAAAAAAAAAGATGAAAGTATTTATCTCTCTTTTGGATTAATTTAAATCA 6170
 QY 61 TTTTTCCTCCCAATGATATATATAAATTTGGATTAATTAATTCGTTT 120
 DB 6171 TTTTTCCTCCCAATGATATATATAAATTTGGATTAATTAATTCGTTT 6230
 QY 121 TTAGTTCGGCTTGAAGAAAGGTTTCGACTTGAAGATGACATGATATAGATTGG 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 6231 TTAGTTCGGCTTGAAGAAAGGTTTCGACTTGAAGATGACATGATATAGATTGG 6290
 QY 181 GACCTAGCTGAGCTCTTGGACATTTGTATGATGTTGATTATAGTGCACACT 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 6291 GACCTAGCTGAGCTCTTGGACATTTGTATGATGTTGATTATAGTGCACACT 6350
 QY 241 ATTAAACCTTAATGGGCTTCTATTAAGGCCCAATTATATAGATTATTAACAAAGTGAC 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 6351 ATTAAACCTTAATGGGCTTCTATTAAGGCCCAATTATATAGATTATTAACAAAGTGAC 6410
 QY 301 AACTTTACTTCGTTTGTGATCCGAAGCATATACAAATGTCACAAATCCAAACGACAG 358
 DB 6411 AACTTTACTTCGTTTGTGATCCGAAGCATATACAAATGTCACAAATCCAAACGACAG 6468
 RESULT 2
 ID ABL33641 standard; DNA: 5449 BP.
 AC ABL33641;
 XX 26-MAR-2002 (first entry)
 DT
 XX Human immune system associated gene SEQ ID NO: 1614.
 DE
 XX Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytosatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX 03-JAN-2002.
 XX 02-JUL-2001; 2001MO-EP07537.
 XX 30-JUN-2000; 2000DE-1032529.
 XX 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 PA Olek A, Plepenbrock C, Berlin K;
 PI WPI: 2002-130909/17.
 DR
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX Claim 1; SEQ ID NO 1614; 32pp + Sequence Listing; German.
 PS
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SO Sequence 5449 BP; 1146 A; 90 C; 1299 G; 2914 T; 0 other;
 Query Match 16.2%; Score 58; DB 24; Length 5449;
 Best Local Similarity 65.4%; Pred. No. 0.053;
 Matches 85; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
 QY 4 AAAAAAAAAAGATGAAAGTATTTATCTCTCTTTTGGATTAATTTAAATCATTT 63
 ||||||||||||||||||||||||||||||||||||||||||||||||||||

	Query Match Similarity	15.8%	Score 56.4;	DB 24;	Length 5926;
	Best Local Similarity	48.2%	Pred. No. 0.1;		
	Matches 159;	Conservative	0;	Mismatches 171;	Indels 0; Gaps 0
OY	4	AAAAAAAAAGANGAAGATATTTTAACTCTCTCTTTTTTGATAATTTTAATCATTTT	63		
Dd	3171	AAAGTAGGGGTTTCAGCTGCATTGTGTTTTTTTATTTTGTGAATTTTATAGAG	3230		
OY	64	TTTTTGCCCAATGATATATAAAAAATTTGGATMAATAATTTATGGATTTGGTTTTTA	123		
Dd	3231	TTTGAATATGCTATTTTATATAGACTTTGTTTGGATTGCGAATTTAGAAAATTAATATTAT	3290		
OY	124	GTTCCGGTTTGAGAAAAGGGTTTCACCTTTCCAAAGTGAGCAGATATATAGATTGGAG	183		
Dd	3291	TATGAGATATTAAGAATTAGTTTATAGGAAGAAGATTTTGGAAATGAATATTTTGGG	3350		
OY	184	CTAGCTTGAGCTTTGGACATTTGTAATTGGAGTGTGGTATTAATAGTGTGACACTAT	243		
Dd	3351	GGAGTTTATAGATATTTATTTATTTGTTGAGCTTATTCGATTAATAATTTTATGTTTAT	3410		
OY	244	AACCTTTAATNGGGCTTCTATTAAGGCCAATTATATATACGATTAATAACAAGTGACAC	303		
Dd	3411	TTAACGAAAGTTGGGTTGGTTGGAGGATTTGGGTAGAGTACGTTGGAAAGTATGTTAT	3470		
OY	304	TTTTACTTCTGTTTTCATCCGACGAATPA	333		
Dd	3471	TATATTTGATTTTGTGAATGTAATTAATPA	3500		
RESULT 5					
AAS46510					
ID	AAS46510	standard:	DNA;	12356	BP.
XX	AAS46510;				
AC	AAS46510;				
DT	18-DEC-2001	(first entry)			
DE	Tumour suppressor gene derived chemically modified sequence #232.				
KW	Human: tumour suppressor gene; oncogene; alitumour; cytostatic;				
KM	cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;				
OS	Cytosine methylation; ds.				
XX	Homo sapiens.				
PN	WO200168912-A2.				
PD	20-SEP-2001.				
PF	15-MAR-2001; 2001WO-EP02955.				
PR	15-MAR-2000; 2000DE-1013847.				
PR	06-APR-2000; 2000DE-1019058.				
PR	07-APR-2000; 2000DE-1019173.				
PR	30-JUN-2000; 2000DE-1032529.				
PA	01-SEP-2000; 2000DE-1043826.				
PI	(EPIG-) EPIGENOMICS AG.				
PI	Olek A, Piepenbrock C, Berlin K;				
DR	WPI; 2001-602752/68.				
PT	Fragments of chemically modified genes associated with tumour suppressor				
PT	genes and oncogenes, useful in designing primers and probes for				
PT	analysing diseases associated with cytosine methylation state e.g.				
PS	cancer				
PS	Claim 1; SEQ ID NO 232; 27pp; English.				
CC	The invention relates to a nucleic acid comprising a sequence of 18				
CC	bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with				

CC	Disruptive, of genes associated with tumour suppression and
CC	oncogenes having a sequence taken from 536 (actually 533 since
CC	numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC	(Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC	peptide nucleic acid-oligomer (PNA) or at least 9 nucleotides and may
CC	form part of a set of probes for detecting the cytosine methylation state
CC	and/or single nucleotide polymorphisms and also to be used in an
CC	array for analysing diseases associated with CpG dinucleotides e.g.
CC	cancers and tumours. The probes can also be used in a method for
CC	ascertaining genetic and/or epigenetic parameters for the diagnosis
CC	and/or therapy of existing diseases or the predisposition to specific
CC	diseases, by analysing cytosine methylations. The parameters may be
CC	compared to another set of genetic and/or epigenetic parameters, the
CC	differences serving as basis for diagnosis and/or prognosis events which
CC	are diagnostic sequences to patients. The present sequence is one of the
CC	533 genomic sequences derived from tumour suppressor genes and
CC	oncogenes. Sequences with even numbered Seq ID numbers are the
CC	complementary sequence of the corresponding odd numbered sequence (e.g.
CC	ID 2 and ID1, ID 536 and ID 533, except for those whose partner sequence
CC	is missing).
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pcl_sequences.
CC	
XX	Sequence 12356 BP; 3645 A; 68 C; 2127 G; 6516 T; 0 other;
XX	
XX	Query Match
XX	Best Local Similarity - 15.5%; Score 55.6; DB 22; Length 12356;
XX	Matches 118; Conservative 0; Mismatches 104; Indels 0; Gaps 0
XX	
QY	11 AGATGAAAGTATTTTATTCCTCTTTTATTTTTCATATTTTAAATCATTTTTCGC 70
DB	9051 AGATATATATGAGTACTTTTATTTTATTTTGTGAGCAATTTTAAATCTTTTATATG 9110
QY	71 CCAATGATATATATTTTAAATTTGATATTAATATATATTCCTTTTATGTCGGG 130
DB	9111 GGTGTATTTGATTTTATTTTATTTTATTTTAAATATATTAAGGCTTTTATTTTATATTTGT 9170
QY	131 TTTGAGAAAAGGGTTTCGACTTTCGAAAGTGCAGCATATATATGATTTGGAGCTAGGTT 190
DB	9171 TATTTAGTATTTTGTATATGTTCTTTTTCGATATTAAGTATTTTAAATGCGTAAATGAT 9230
QY	191 GAGCTCTTGACATTTGATTTGATGCTGTTATATTTAGTC 232
DB	9231 ATTATATTTGATTTGATTTTGTATTTTTCATGATTTATG 9272
XX	
XX	RESULT 6
XX	ABL32028
XX	ABL32028 standard; DNA: 7351 BP.
XX	ABL32028;
XX	
XX	26-MAR-2002 (first entry)
XX	
DE	Human immune system associated gene SEQ ID NO: 1.
XX	
KX	Human; immune system disease; cytosine methylation; antislammatic;
KX	antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KX	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KX	antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KX	antiflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KX	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KX	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KX	gene; ds.
OS	Homo sapiens.
XX	
XX	MO200200928-A2.
XX	
XX	03-JAN-2002.
XX	

KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KV	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KX	gene; ds.
OS	Homo sapiens.
XX	
PN	MO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PE	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A. Piepenbrock C, Berlin K;
DR	WPI; 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal
PS	cytosine methylation -
XX	
PS	Claim 1; SEQ ID NO 2100; 32pp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC	rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC	diseases. The present sequence is a gene of the invention.
XX	
SQ	Sequence 2814 BP: 629 A; 144 C; 770 G; 1271 T; 0 other;
	Query Match 15.1%; Score 54.2; DB 24; Length 2814;
	Best Local Similarity 48.8%; Pred. No. 0.26;
	Matches 146; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
QY	20 TATTTTATTCCTCTTTTTTTTTTGATAATTAAATCATTTTTTGGCCCAATGATA 79
DB	3 TTTGTTAATTATTAATTATTTATTTATTAATAAATAAATAAATAAATAAATTATATA 62
QY	80 TATAAAAAATTTGGATAAAATAATATTTAGATATTCGTTTTAGTCGGTTTGAGAAA 139
DB	63 TAGCGTAGGAGGATATATATTTTGTATATATTTTCGTATATAGTAATTTTGAT 122
QY	140 AGCGTTTCGACTTTCGAAGAATGCACGATATATAGATTGGAGCTAGCTTACGCTTG 199
DB	123 ATGATATGATTAATATATATATATGTCGATATAGCTTTTTTGTATAAACGTTTATA 182
QY	200 GACATTTTGTATGGATGCTGTTGATTTATAGTCGACACATTTAAACCTTAATGGCT 259
DB	183 TGTAATATATATATTAATATTTTATATATATATAGTTATATTTAGTATTTGCGT 242
QY	260 TTCTATTAAGGCCAATTAATATAGCATTAATAACAAGTGACAAGCTTTACTTCGTTTT 318
DB	243 TTTGGAGATTTTAAATTTTATTTTATAGATTAATAAGCAGAGATTTTATTTT 301
RESULT 9	
ABR28455	
ID	ABR28455 standard; DNA: 7138 BP.
AC	ABR28455;
DT	23-APR-2002 (first entry)
XX	
XX	DNA transcription associated genomic DNA #165.

KW	DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KV	PNA: cytosine methylation state; SNP: retroviral infection; gene; ds;
KM	single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KM	vital infection; Seary syndrome; haematological disorder; tuberculosis;
KM	immunological disorder; Werner syndrome; developmental disorder;
KM	psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KM	neurodegenerative disorder; Maardenburg syndrome; Niemann-Pick disease;
KW	myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW	polyglutamine disorder; congenital heart disease; HDR syndrome; gene therapy;
XX	
OS	Unidentified.
XX	
PM	WO200192565-A2.
PD	
XX	
XX	06-DEC-2001.
PF	
XX	
PR	06-APR-2001; 2001WO-EP03973.
PR	06-APR-2000; 2000DE-1019058.
PR	07-APR-2000; 2000DE-1019173.
PR	30-JUN-2000; 2000DE-1033529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EPIG-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
DR	
XX	WPI; 2002-090046/12.
XX	
PT	New nucleic acids or oligomers, useful for diagnosing or treating
PT	diseases associated with DNA transcription, e.g. immunological
PT	disorders, Werner syndrome, psoriasis, myocardial infarction, solid
XX	tumours or cancer -
PS	
XX	Claim 1; SEQ ID No 329; 32pp; English.
XX	
CC	The invention relates to a nucleic acid, which comprises a segment of the
CC	chemically pretreated DNA of genes associated with DNA transcription from
CC	one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC	or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC	to the chemically pretreated DNA of genes associated with DNA
CC	transcription. The set of oligomer probes are useful for detecting the
CC	cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC	in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC	diagnosing or treating diseases associated with DNA transcription
CC	(particularly with the methylation status), e.g. adenosine deaminase
CC	deficiency, viral infection, retroviral infection, seary syndrome,
CC	haematological disorders, immunological disorders, Werner syndrome,
CC	tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC	neurological disorders, neurodegenerative disorders, Maardenburg
CC	syndrome, Mismann-Pick disease, myelodysplastic syndrome, myocardial
CC	infarction, hypertension, angiodysgenesis, erythropoiesis, congenital heart
CC	disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC	or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC	associated genomic DNA molecules of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification but was obtained in electronic format directly from the
CC	European Patent Office.
XX	
SQ	Sequence 7138 BP; 1564 A; 288 C; 1759 G; 3527 T; 0 other:
	Query Match 15.1%; Score 54; DB 24; Length 7138;
	Best Local Similarity 48.1%; Pred. No. 0.27;
	Matches 153; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
OY	3 AAAAAAAAAAGATGAAGTATTTTTATTCCTCTCTTTTTTTTGATTAATTTAAATCAT 62
Db	3163 AAAAAAATATAGTAGACGGCGGTGTTATTTTATTTTATTTTGGTATTTGCATTTGCT 3222
OY	63 TTTTTTGGCCAATGATATATAAAAATTTCGATAAATATATATATTCGATTCGTTTTTT 122
Db	3223 TTTTGGACATATGTTATTTATTTTATTTTATTTTATTTTATTTTATGTTTGGTTTTTTTT 3282

QY 123 AGTCGGGTTTGAGAAAGGCTTTCACATTCGAAAGTGACATGATATGATTGGCA 182
DB 3283 AATTAGCTTTTAAAGATGATTTTGTATATAGTATGCTTTTATATGATGCTAA 3342
QY 183 GGTAGGTTGAGCTTTGACATTTGATTTGAGTGTGATTTATGATGACACTAT 242
DB 3343 AATATTTTATTTTATTTATTTATTTGATTTTATTTATTTATTTGCGGTAA 3402
QY 243 TAAACCTTAATGCGCTTCTATAGGCCCAATATATTCAGATTTTAAACAAGTGACAA 302
DB 3403 TTAAAGTATATTTTATTTATTTAGTGTGATTTTGTGATTTTATTTATTTAGAA 3462
QY 303 CTTTACTTCGTTTTTGA 320
DB 3463 ATTTTATAGTTTTTTAA 3480

RESULT 10
AAS46743
ID AAS46743 standard; DNA; 14147 BP.
AC AAS46743;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #467.
XX
KW Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; Cpg dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
XX
OS
XX
PS Claim 1: SEQ ID No 467; 27pp: English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (SS) and sequences complementary to (SS). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with Cpg dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be

CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp://ipo.int/published_sequences.
XX
SQ Sequence 14147 BP; 3621 A; 225 C; 3116 G; 7185 T; 0 other;
Query Match 14.9%; Score 53.2; DB 22; Length 14147;
Best Local Similarity 48.1%; Pred. No. 0.36;
Matches 151; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
QY 7 AAAAAAGTGAAGATTTTATCTCTCTTTTATTTTATTTTAAATGATTTT 66
DB 10811 AAAAAAGTGTGTTAGTGTGTTGCTTTGTTATTTTGTGTTTTCGTTGTTT 10870
QY 67 TTGCCCAATGATATATTAATTTGATTAATATTTATTTGATTTTCGTTTATGTT 126
DB 10871 TTTTTCGAATTTATGTTGTTTATTTTATTTATTTGTTAGTTATTTTATGTT 10930
QY 127 CGGCTTGAGAAAAGGTTTCGACTTTCGAAAGTGACGATATATTTGGACCTA 186
DB 10931 TTGGGATTTTATATTCGTTGCTGTTTATTTGTTGTTGTTTATTTAGATT 10990
QY 187 GGTGAGCTTTTGACATTTGATTTGATTTGTTGTTATTTAGTTCGACATTTAA 246
DB 10991 TTTAGTTTATTTTATTTTATTTGTTTATTTGTTTATTTTGGCGATTTATTT 11050
QY 247 CCTTAATGCGCTTCTATAGGCCCAATATATTTAGTATTAATTAACAAGTGACAACTTT 306
DB 11051 TTTTTCGATTTTATGTAATTAAGTATTTTATTTTGTGTTGATTTTATTAATTT 11110
QY 307 TACTTCGTTTTTGA 320
DB 11111 TTATTTATTTATTTGA 11124

RESULT 11
ABK33955
ID ABK33955 standard; DNA; 14147 BP.
AC ABK33955;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human DNA for staging of Astrocytomas #20.
XX
KW Human: ds; astrocytoma; cytostatic; staging; cysteine methylation; Cpg;
KW bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;
KW matrix assisted laser desorption/ionization mass spectrometry.
XX
OS Homo sapiens.
XX
PN WO200202808-A2.
XX
PD 10-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07538.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-171649/22.
XX
PT Novel chemically modified genomic DNA sequences, useful in the

CC specification, but was obtained in electronic format directly from WIPPO
at ftp.wippo.int/pub/published_pcr_sequences.

Sequence 21231 BP; 6727 A; 3859 C; 3768 G; 6877 T; 0 other;

Query Match 14.7%; Score 52.6; DB 23; Length 21231;

Best Local Similarity 46.8%; Pred. No. 0.46; Mismatches 189; Indels 0; Gaps 0;

Matches 166; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

1 AAAAAAAAAAGATGAAAGTATTTATCTCTCTTTTGTGAAATTTTAAATCA 60

16810 AAAAGTCAGAGTCAATTAATAAAGCTGTGATTTTATATTTTATATAT 16869

61 TTTTTCCTCCCAATGATATATAAATTTGATTAATATATTTGATTTGCTTT 120

16870 AATATATATAGAAATATATATATTTATTTAAACAAATGCTATATATATCAGTTT 16929

121 TTAGTTCGGCTTTGAGAAAAGGTTTCACCTTCGAAAGTGACGATATATAGCTTG 180

16930 GAAAAAATTTTAAAGTCTTTTATTTATTTGAAAAATGATCTTAAATATTTCT 16989

181 GAGCTAGGTGAGTCTTTGACATTTGATTTGATTTGATTTGATTTGATTTG 240

16990 AATATATTTTTCACCTTTGTAATATGATCTGATTTATATAAAGCTATTTTGG 17049

241 AATTAACCTTAATGAGCTTTCTATATAGGCCCAATTAATTTGATTTATACAAAGTGC 300

17050 GCCAAGCCAAAGTGTCTGCTTTAGCAATCGAAAAATTTTACGATTTTGAAGAAAG 17109

301 AACTTTTACTGCTTTTGTGATCGCAAGCAATTAACAATTTGCAATTAACCAACAC 355

17110 CATCGATTTTCTATTTTCAATGGAATTTGTAAATCTTACGAATTCGATTTAAC 17164

RESULT 15

ABL32322

ID ABL32322 standard; DNA; 6418 BP.

AC ABL32322;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 295.

XX Human; immune system disease; cytosine methylation; antiasthmatic;

KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antirheumatic; antidiabetic; antiparasitic;

KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KW gene; ds.

OS Homo sapiens.

XX WO200200928-A2.

PD 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

PA Olek A. Phippenbrock C. Berlin K;

PI WPI; 2002-130909/17.

DR Nucleic acid comprising fragment of chemically modified gene, useful

XX for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation

XX Claim 1; SEQ ID NO 295; 32pp + Sequence Listing; German.

PS The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX Sequence 6418 BP; 1842 A; 47 C; 1423 G; 3106 T; 0 other;

QY Query Match 14.6%; Score 52.2; DB 24; Length 6418;

Best Local Similarity 56.8%; Pred. No. 0.57;

Matches 96; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

18 AGTATTTTATCTCTCTTTTGTGATATTTTAAATCAATTTTGTGCGGATGAGA 77

1110 AATTTTATTTATAGTCTTTTGTGATTTTATTTTATTTTATTTTGTGCTAGTTG 1169

78 TATATTAATTTGATTAATATATTTGATTTGATTTGATTTGATTTGATTTG 137

1170 TATTTTATTTGATTTTATTTTAAATATTAATTTTGTGTTTGTGATTAAGTA 1229

QY 138 AAGGTTTCGACTTTCGAAAGTGACGATATATATGATTTGGAGCTA 186

1230 TATTTGATTTATTTTGTGAGGATTAATTTATGTTTATTTATTTA 1278

Search completed: March 29, 2003, 19:00:18
Job time : 138.508 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:09:40 ; Search time 12.5196 Seconds
(without alignments)
8769,450 Million cell updates/sec

Title: US-09-502-426A-1_COPY_6111_6468

Perfect score: 358
Sequence: 1 aaaaaaaaaagatgaagtc.....tgtcaatatacaaacacaag 358

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/lna/PCRTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/lna/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42.8	12.0	1493	1	US-08-340-820-24
2	42.8	12.0	1493	1	US-08-593-535-24
3	41.8	11.7	998	4	US-09-122-400B-5
4	41	11.5	2230	4	US-08-378-313-24
5	40.8	11.4	6243	2	US-09-056-075-1
6	40.8	11.4	19557	5	PCRT-US92-06300-1
7	40.6	11.3	1733	3	US-09-073-569-1
8	40.6	11.3	3095	6	5231168-1
9	40.4	11.3	10607	1	US-08-078-090-3
10	40	11.2	2755	3	US-08-749-522-2
11	39.4	11.0	289	1	US-08-341-568-3
12	39.4	11.0	289	2	US-08-911-020-3
13	39.4	11.0	19124	2	US-08-487-826B-13
14	39	10.9	1895	4	US-09-444-336-7
15	39	10.9	4285	4	US-09-410-464-1
16	39	10.9	9048	3	US-08-973-273-4
17	38.8	10.8	835	4	US-08-998-416-547
18	38.8	10.8	863	4	US-08-998-416-498
19	38.8	10.8	3138	1	US-07-867-106-4
20	38.2	10.7	2836	3	US-08-747-221B-24
21	38.2	10.7	2836	3	US-08-747-221B-26
22	38.2	10.7	2836	4	US-09-005-051-24
23	38.2	10.7	2836	4	US-09-005-051-26
24	38.2	10.7	5852	1	US-07-867-106-2
25	38.2	10.7	29604	3	US-08-781-891-207
26	38	10.6	552	4	US-09-134-001C-2606
27	38	10.6	701	4	US-08-998-416-701

28	38	10.6	2555	2	US-08-693-457-3	Sequence 3, Appl
29	38	10.6	2555	4	US-09-265-731-3	Sequence 3, Appl
30	38	10.6	8700	2	US-08-392-625-16	Sequence 16, Appl
31	38	10.6	8700	2	US-08-466-961A-16	Sequence 16, Appl
32	38	10.6	8700	2	US-08-645-193B-18	Sequence 18, Appl
33	37.8	10.6	1582	3	US-08-845-196B-10	Sequence 10, Appl
34	37.8	10.6	1582	3	US-08-845-196B-12	Sequence 12, Appl
35	37.6	10.5	2107	4	US-09-180-852-1	Sequence 1, Appl
36	37.6	10.5	2422	1	US-07-867-106-5	Sequence 5, Appl
37	37.6	10.5	3138	1	US-07-867-106-4	Sequence 4, Appl
38	37.6	10.5	5852	1	US-07-867-106-2	Sequence 2, Appl
39	37.6	10.5	6078	4	US-09-173-914-1	Sequence 1, Appl
40	37.4	10.4	477	4	US-08-887-534A-81	Sequence 81, Appl
41	37.4	10.4	2445	6	5215909-9	Patent No. 5215909
42	37.4	10.4	2674	4	US-09-817-180-1	Sequence 1, Appl
43	37.4	10.4	3761	4	US-08-890-865A-2	Sequence 2, Appl
44	37.4	10.4	10223	4	US-08-961-527-73	Sequence 73, Appl
45	37.2	10.4	144	1	US-08-702-344-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-340-820-24/c
Sequence 24, Application US/08340820
Patent No. 5512460
GENERAL INFORMATION:
APPLICANT: NARUO, Ken-ichi
APPLICANT: SEKO, Chisako
APPLICANT: KUROKAWA, Tsutomu
APPLICANT: KUNDO, Tatsuya
TITLE OF INVENTION: GLIA ACTIVATING FACTOR AND ITS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
ADDRESSEE: CUSHMAN
STREET: 130 Water Street
City: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,820
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/835,713
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: CONLIN, David G.
REGISTRATION NUMBER: 27026
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

Db 1116 TTTTACTTACTTACTTACTTCCGAGGAA 1089

RESULT 6

PCT-US92-06300-1/c

Sequence 1, Application PC/TUS9206300

GENERAL INFORMATION:

APPLICANT: Hurwitz, David R

APPLICANT: Nathan, Margaret

APPLICANT: Shanl, Moshe

TITLE OF INVENTION: Transgenic Protein Production

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Rhone-Poulenc Rorer, Inc.

STREET: 500 Virginia Ave., Bldg. 3A

CITY: Ft. Washington

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19034

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/06300

FILING DATE: 1992/07/30

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Goodman, Rosanne

REGISTRATION NUMBER: 52,534

REFERENCE/DOCKET NUMBER: A0856-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 962-4107

TELEFAX: (215) 962-4130

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1957 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOHETICAL: NO

ANTI-SENSE: NO

PUBLICATION INFORMATION:

AUTHORS: Minghetti, P P

AUTHORS: Ruffner, D E

AUTHORS: Kuang, W.-J

AUTHORS: Dennis, O E

AUTHORS: Hawkins, J W

AUTHORS: Beattie, W G

AUTHORS: Dugalczyk, A

TITLE: MOLECULAR STRUCTURE OF THE HUMAN ALBUMIN GENE

TITLE: IS REVEALED BY NUCLEOTIDE SEQUENCE WITHIN Q11-22

JOURNAL: J. Biol. Chem.

VOLUME: 261

PAGES: 6747-6757

DATE: 1986

RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 19002

PCT-US92-06300-1

Query Match 11.4%; Score 40.8; DB 5; Length 19557;

Best Local Similarity 52.3%; Pred. No. 2.3;

Matches 90; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Db 2 AAAAAAAAAAGTGAAGTATTTTATCTCTCTTTTGTGATAATTTAAATCAT 61

Db 423 ATATCATATATATGATATATTTCAATTTCTTTTGTGATAATTTAAATCAT 364

Db 62 TTTTGTGCGCAATGATATATTTGATTAATATATATGATATCTGTTT 121

Db 363 TTTTGTGCGCAATGATATATTTGATTAATATATGATATCTGTTT 304

Db 122 TACTGCGGTTTGAGAAAGGTTTCGATTCGAAAGTGAGAGTATAT 173

Db 303 TTTGTTGTTTGTGTTTATTTTGTGTTTATATATCTGTCGTGTCGT 252

RESULT 7

US-09-073-569-1/c

Sequence 1, Application US/09073569

Patent No. 6084088

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Grossmann, Angelika

TITLE OF INVENTION: NOVEL TUMOR ANTIGENS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Zymogenetics, Inc.

STREET: 1201 Eastlake Avenue East

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98102

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/073,569

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sawislak, Deborah A

REGISTRATION NUMBER: 37,438

REFERENCE/DOCKET NUMBER: 97-14

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-442-6672

TELEFAX: 206-442-6678

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1733 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 34...1344

OTHER INFORMATION:

US-09-073-569-1

Query Match 11.3%; Score 40.6; DB 3; Length 1733;

Best Local Similarity 56.3%; Pred. No. 2.5;

Matches 76; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Db 20 TATTTTATCTCTCTCTTTTGTGATAATTTAAATCATTTTGTGCGCAATGATA 79

Db 1715 TTTTGTGCGCAATGATATATTTGATTAATATATGATATCTGTTT 1656

Db 80 TTTTAAATTTTGATTAATATTTTGTGATAATTTGATTAATTTTGTGCGCAATGATA 139

Db 1655 TTTTGTGCGCAATGATATATTTGATTAATATATGATATCTGTTT 1596

Db 140 AGGTTTGCATTTTC 154

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

Db 1595 AGGTTTGCATTTTC 1581

5231168-1
; Patent No. 5231168
; APPLICANT: DZIEGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; SEQ ID NO: 1:
; FILING DATE: 18-SEP-1989
; LENGTH: 3095
5231168-1
Query Match 11.3%; Score 40.6; DB 6; Length 3095;
Best Local Similarity 62.1%; Pred. No. 2.5;
Matches 64; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
OY 4 AAAAAAAAAAGATGATATTTATCTCTCTTTTGGATTAATTTAAATTTT 63
DB 2724 AAGAAGAGAAAAAATTTTATATATCAATTTTCAATTAATTAATTAATTA 2783
OY 64 TTTTGGCCATGATATATTAATAATTTGGATTAATTAATTAAT 106
DB 2784 TTATATATAAATATCGATTAATTTATGATATATTAATTAAT 2826
RESULT 9
US-08-078-090-3
; Sequence 3, Application US/08078090
; Patent No. 5739407
; GENERAL INFORMATION:
; APPLICANT: BERGSTROM, SVEN
; APPLICANT: HERNELL, OLLE
; APPLICANT: LOENNERDAL, BO
; APPLICANT: HJALMARSSON, KARIN
; APPLICANT: HANSSON, LENNART
; APPLICANT: TOERNELL, JAN
; APPLICANT: STROMQUIST, MATS
; TITLE OF INVENTION: HUMAN BETA-CASEIN PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 SEVENTH STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/078, 090
; FILING DATE: 19930618
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NO PCT/DK92/00236
; FILING DATE: 19-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NO PCT/DK91/00233
; FILING DATE: 19-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: COOPER, IVER P.
; REGISTRATION NUMBER: 28, 005
; REFERENCE/DOCKET NUMBER: BERGSTROM2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)628-5197
; TELEFAX: (202)737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:

LENGTH: 10607 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(4804..4854, 5720..5746, 6726..6746, 6845
; ..6886, 7991..8521, 9440..9445)
US-08-078-090-3

Query Match 11.3%; Score 40.4; DB 1; Length 10607;
Best Local Similarity 56.9%; Pred. No. 2.8;
Matches 74; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY 60 ATTTTGGCCCATGATATATAAATTTGGATTAATTAATTAATTTGATTCGTT 119
DB 923 ATTATAGTATAGAAATAGATATAAAGCTGATTAATTAATTAATTAATTAAT 982
OY 120 TTTAGTTGGGTTTGAGAAAGGTTTCGACTTCGAAAGTGACGATATATAGATTG 179
DB 983 ATTACTATATTAATTAATGATGATTAATTAATTAATTAATTAATTAATTAATTA 1042
OY 180 GGACCTAGCT 189
DB 1043 GTAGATTAGT 1052

RESULT 10
US-08-749-522-2
; Sequence 2, Application US/08749522
; Patent No. 6096950

; GENERAL INFORMATION:
; APPLICANT: John, Mallyakal
; TITLE OF INVENTION: FIBER-SPECIFIC PROMOTERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: U.S.A.
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,522
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 670513.90244
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-749-522-2

Query Match 11.2%; Score 40; DB 3; Length 2755;
Best Local Similarity 48.3%; Pred. No. 3.3;
Matches 171; Conservative 0; Mismatches 180; Indels 3; Gaps 2;
OY 6 AAAAAAAAAAGATATTTATCTCTCTTTTGGATTAATTTAAATCATTTT 65

; Sequence 13, Application US/08487826B
; Patent No. 5993827

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION: 435
NAME: Israel, Ned
REGISTRATION NUMBER: 29, 655
REFERENCE/DOCKET NUMBER: NH121.001CPL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEO ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
JS-08-487-826B-13

Best Local Similarity 46.7%; Pred. No. 4.4; Mismatches 161; Conservative 0; Mismatches 181; Indels 3; Gaps 1

QY	6	AAAAAAGTCAAGAGTATTTTATTCGTCCTTTTATTTTATTAATTTTAAATCATATTTT	65
Db	15815	AAATTAATTTTATTTTATTTATGATATATTTTATTTTAAACATTATTTTAAATTTT	15755
QY	6	TTTGCCCATGATATATTAATAAATTTGGATAAATAATATATATGATATTCGCTTTTACT	125
Db	15755	TATTTATGATATATATATTTTATTTTATATATTTTCTTTTATTTTGTTTTATGA	15666
QY	126	TGCGGTTGAGAAAAGGGTTTCGACTTTCGAAAGTGCATGATATATA--GATGGGA	182
Db	15695	TATATATTTTATTTTATTTTAAAGTTTTTTTCTCTCTTGTTTATTTTATTTTATA	15636

[illegible]

US-09-444-336-7/c
; Sequence 7, Application US/09444336
; Patent No. 6410713

```

: GENERAL INFORMATION:
: APPLICANT: Guerriero, Vincent
: APPLICANT: Raynes, Deborah A
: TITLE OF INVENTION: DNA Encoding Proteins That Inhibit Hsp70 Function
: FILE REFERENCE: HspB DNA and Protein Sequences
: CURRENT APPLICATION NUMBER: US/09/444,336
: CURRENT FILING DATE: 1999-11-19
: EARLIER APPLICATION NUMBER: 60/109,351
: EARLIER FILING DATE: 1998-11-20
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 7
: LENGTH: 1895
: TYPE: DNA
: ORGANISM: Brachydanio rerio (zebra fish)
US-09-444-336-7

```

Best Local Similarity 63.28; Pred. No. 5.2;
Matches 60; Conservative 0; Mismatches 35; Indels

QY	20	TATTTTATCTCTCTCTTTTTTTTGATAATTAAATCAATTTTTTTGCCAATGATA	79
Db	1862	TTTGTAAACACA	1803
QY	80	TATAAAAATTGGATPAATAATATATTGATATT	114
Db	1802	GGCTGATGATGTATTATTTTAAATATGCTATTT	1768

Sequence 1, Application US/09410464
Patent No. 6395892

```

: GENERAL INFORMATION:
: APPLICANT: Strauss et al.
: TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in
: TITLE OF INVENTION: poplar and other plant species.
: FILE REFERENCE: 53375
: CURRENT APPLICATION NUMBER: US/09/410,464
: CURRENT FILING DATE: 1999-10-01
: EARLIER APPLICATION NUMBER: 09/287,700
: EARLIER FILING DATE: 1999-04-06
: EARLIER APPLICATION NUMBER: 60/080,851
: EARLIER FILING DATE: 1998-04-06
: NUMBER OF SEQ ID NOS: 24
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 4285
: TYPE: DNA
: ORGANISM: Populus balsamifera subsp. trichocarpa
: US-09-410-464-1

```

Best Local Similarity	55.68;	Pred. No. 5.2;	
Matches	75; Conservative	0;	Mismatches
			60; Indels
			0; Gaps


```
Publication No. US20030054421A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA103P1C1
CURRENT APPLICATION NUMBER: US/10/102,806
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: 09/925,298
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05881
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 846
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 401
LENGTH: 189
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (11)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (162)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (165)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (166)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (187)
OTHER INFORMATION: n equals a,t,g, or c
US-10-102-806-401

Query Match          12.6%; Score 45; DB 9; Length 189;
Best Local Similarity 63.3%; Pred. No. 2.1;
Matches 69; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

OY 20 TATTTTATCTCTCTCTTTTGGATTAATTTTAAATTTTGGCCCAATGATA 79
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 158 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 99
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 80 TATTAATAATTTGGATAAATATATATATGATATCGTTTGTAGTCG 128
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 98 TATTAATAATTAACATGTATGACAGGTTTCAACAATTTGCTGTAGTTTG 50

RESULT 3
US-09-960-352-12673/c
Sequence 12673, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 12673
LENGTH: 277
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 54-LIB3058-039-Q1-K1-F10
```

```
US-09-960-352-12673
Query Match          12.5%; Score 44.6; DB 10; Length 277;
Best Local Similarity 57.6%; Pred. No. 2.8;
Matches 80; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 2 AAAAAAAAAAGATGAAAGTATTTTATCTCTCTTTTGGATTAATTTTAAATCAT 61
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 275 AAATTTAATAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTCA 216
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 62 TTTTTCGCCCAATGATATATTAATAATTTGATTAATATATGATATTCGTTT 121
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 215 AAATTTTTCATTTTAAAAATTTTCTTTTAAATTTTAAATTTTAAATTTT 156
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 122 TAGTTCGGGTTTGAGAAA 140
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 155 TTTTCTTTTCTTTTAAAAAAA 137
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
US-09-822-830A-26
Sequence 26, Application US/09822830A
Patent No. US20020142952A1
GENERAL INFORMATION:
APPLICANT: Genetics Institute, Inc.
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakari
APPLICANT: Graham, James R.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6402
CURRENT APPLICATION NUMBER: US/09/822,830A
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195,604
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 631
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 681
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-830A-26

Query Match          12.5%; Score 44.6; DB 10; Length 681;
Best Local Similarity 53.1%; Pred. No. 3.6;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 20 TATTTTATCTCTCTCTTTTGGATTAATTTTAAATTTTGGCCCAATGATA 79
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 321 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 380
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 80 TATTAATAATTTGGATAAATATATATGATATTCGTTTGTAGTCGAGTTGAGAAA 139
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 381 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 440
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 140 AGGCTTTCACCTTCGAAGTGACGATGATATATGAGTGGAGCTAGTGCTTT 198
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 441 ATTAAGCTTATTAACATCAAGTAAATATACATCAAGTAAAGTCAACAGTTTGTGCTTT 499
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
US-09-960-352-5094
Sequence 5094, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
```


1 PRIOR APPLICATION NUMBER: 60/0853233
2
3 PRIOR FILING DATE: 1998-05-13
4
5 PRIOR APPLICATION NUMBER: 60/085338
6
7 PRIOR FILING DATE: 1998-05-13
8
9 PRIOR APPLICATION NUMBER: 60/085339
10
11 PRIOR FILING DATE: 1998-05-13
12
13 PRIOR APPLICATION NUMBER: 60/085579
14
15 PRIOR FILING DATE: 1998-05-15
16
17 PRIOR APPLICATION NUMBER: 60/085697
18
19 PRIOR FILING DATE: 1998-05-15
20
21 PRIOR APPLICATION NUMBER: 60/085704
22
23 PRIOR FILING DATE: 1998-05-15
24
25 PRIOR APPLICATION NUMBER: 60/086414
26
27 PRIOR FILING DATE: 1998-05-22
28
29 PRIOR APPLICATION NUMBER: 60/086430
30
31 PRIOR FILING DATE: 1998-05-22
32
33 PRIOR APPLICATION NUMBER: 60/087106
34
35 PRIOR FILING DATE: 1998-05-28
36
37 PRIOR APPLICATION NUMBER: 60/088026
38
39 PRIOR FILING DATE: 1998-06-04
40
41 PRIOR APPLICATION NUMBER: 60/088730
42
43 PRIOR FILING DATE: 1998-06-10
44
45 PRIOR APPLICATION NUMBER: 60/088741
46
47 PRIOR FILING DATE: 1998-06-10
48
49 PRIOR APPLICATION NUMBER: 60/088810
50
51 PRIOR FILING DATE: 1998-06-10
52
53 PRIOR APPLICATION NUMBER: 60/088858
54
55 PRIOR FILING DATE: 1998-06-11
56
57 PRIOR APPLICATION NUMBER: 60/089532
58
59 PRIOR FILING DATE: 1998-06-17
60
61 PRIOR APPLICATION NUMBER: 60/089559
62
63 PRIOR FILING DATE: 1998-06-17
64
65 PRIOR APPLICATION NUMBER: 60/089907
66
67 PRIOR FILING DATE: 1998-06-18
68
69 PRIOR APPLICATION NUMBER: 60/089947
70
71 PRIOR FILING DATE: 1998-06-19
72
73 PRIOR APPLICATION NUMBER: 60/090349
74
75 PRIOR FILING DATE: 1998-06-23
76
77 PRIOR APPLICATION NUMBER: 60/090429
78
79 PRIOR FILING DATE: 1998-06-24
80
81 PRIOR APPLICATION NUMBER: 60/090445
82
83 PRIOR FILING DATE: 1998-06-24
84
85 PRIOR APPLICATION NUMBER: 60/090538
86
87 PRIOR FILING DATE: 1998-06-24
88
89 PRIOR APPLICATION NUMBER: 60/090863
90
91 PRIOR FILING DATE: 1998-06-26
92
93 PRIOR APPLICATION NUMBER: 60/091360
94
95 PRIOR FILING DATE: 1998-07-01
96
97 PRIOR APPLICATION NUMBER: 60/091519
98
99 PRIOR FILING DATE: 1998-07-02
100
101 PRIOR APPLICATION NUMBER: 60/091982
102
103 PRIOR FILING DATE: 1998-07-07

Query Match	12.18;	Score 43.2;	DB 9;	Length 2933;
Best Local Similarity	63.58;	Pred. No. 10;		
Matches	66;	Conservative		

0;	mismatches	38;	Indels	0;	Gaps	0;
----	------------	-----	--------	----	------	----

[illegible][illegible]

80 TATAAAATTGGATAAATATATATATGGATATTCGTTTTTAA 123

[illegible]

.....CATTI 2822

RESULT 11
ITS-10-121

; Sequence 345, Application NS/10121046

Publication No. US20030022239A1

APPLICANT: Baker, Kevin P.

APPLICANT: Beresini, Maureen

ORGANISM: Homo Sapien
US-10-123-904-345
Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C17
CURRENT APPLICATION NUMBER: US/10/121,049
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-121-049-345

Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C17
CURRENT APPLICATION NUMBER: US/10/121,049
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-121-049-345

RESULT 12
US-10-123-904-345/C
Sequence 345, Application US/10123904
Publication No. US20030022326A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C154
CURRENT APPLICATION NUMBER: US/10/123,904
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA

ORGANISM: Homo Sapien
US-10-123-904-345

Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C17
CURRENT APPLICATION NUMBER: US/10/121,049
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-121-049-345

RESULT 13
US-10-140-470-345/C
Sequence 345, Application US/10140470
Publication No. US20030022331A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C160
CURRENT APPLICATION NUMBER: US/10/140,470
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-470-345

Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C160
CURRENT APPLICATION NUMBER: US/10/140,470
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-470-345

Query Match 12.1%; Score 43.2; DB 9; Length 2933;
Best Local Similarity 63.5%; Pred. No. 10;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C160
CURRENT APPLICATION NUMBER: US/10/140,470
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-470-345

RESULT 14
US-10-175-746-345/C
Sequence 345, Application US/10175746
Publication No. US20030027270A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C160
CURRENT APPLICATION NUMBER: US/10/140,470
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 345
LENGTH: 2933
TYPE: DNA
ORGANISM: Homo Sapien
US-10-140-470-345

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 18:08:52 ; Search time 467.144 Seconds

(without alignments)
12411.569 Million cell updates/sec

Title: US-09-502-426a-1_COPY_6111_6468

Sequence: 358
1 aaaaaaaaaagatgaagt.....tgcataaccacacacaaag 358

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vtc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137.4	38.4	176	17	AL762543 Arabidops
2	128.4	35.9	194	17	AL762544 Arabidops
3	121	33.8	423	10	AV787618 Arabidops
4	68	19.0	1101	17	CNS000TF2 Arabidops
5	65.4	18.3	928	17	CNS00DKY Arabidops
6	60	16.8	905	17	CNS00KHX Arabidops

C	7	58.2	16.3	1027	17	CNS02T50	AL212733 Tetradon
C	8	58	16.2	1101	17	CNS016LI	AL106896 Drosophila
C	9	57.8	16.1	961	17	AZ691838	AL2691838 EMTM156TF
C	10	57	15.9	1101	17	CNS00EJ4	AL0692527 Drosophila
C	11	56.6	15.8	894	17	CNS018BG	AL109126 Drosophila
C	12	56.6	15.8	1147	17	CNS073CX	AL427351 clone BAO
C	13	56.4	15.8	1101	17	CNS0029N	AL097397 Drosophila
C	14	56.2	15.7	1101	17	CNS0021J	AL061936 Drosophila
C	15	56	15.6	949	17	CNS04AIH	AL281906 Tetradon
C	16	55.8	15.6	767	17	CNS00A0X	AL175696 Tetradon
C	17	55.6	15.5	1092	17	CNS020K7	AL419462 r3 end of
C	18	55	15.4	1007	17	CNS000E7	AL065973 Drosophila
C	19	54.8	15.3	1101	17	CNS06KX5	AL065967 Drosophila
C	20	54.6	15.3	952	17	CNS000IEB	AL108927 Drosophila
C	21	54.6	15.3	1002	17	CNS0185X	AL106197 Drosophila
C	22	54.6	15.3	1121	17	CNS01623	AL075520 Drosophila
C	23	54.4	15.2	423	17	CNS001W9	AL099876 Drosophila
C	24	54.2	15.1	866	17	CNS0216T	AL069706 Drosophila
C	25	54.2	15.1	1101	17	CNS00EVL	AL107122 Drosophila
C	26	54	15.1	952	17	CNS016RS	AL561361 AL561361
C	27	54	15.1	993	9	AL561361	AL106578 Drosophila
C	28	54	15.1	1200	17	CNS016CO	AL103735 Drosophila
C	29	53.6	15.0	1043	17	CNS0145P	AL101513 Drosophila
C	30	53.6	15.0	1094	17	CNS012R2	AL427102 clone BAO
C	31	53.6	15.0	1167	17	CNS07360	AL425064 clone BAO
C	32	53.4	14.9	973	17	CNS071LE	AG043196 Pan trogl
C	33	53.4	14.9	974	17	AG043196	AL514085 AL514085
C	34	53.2	14.9	458	9	AL514085	AL057169 Drosophila
C	35	53.2	14.9	994	17	CNS0058N	AL046635 TF end of
C	36	53	14.8	573	17	CNS06LTX	AL237857 Tetradon
C	37	53	14.8	813	17	CNS036TC	AL229845 Tetradon
C	38	52.8	14.7	637	17	CNS036CC	AL106171 Drosophila
C	39	52.6	14.7	1225	9	CNS016ID	AL514657 AL514657
C	40	52.4	14.6	529	9	AL514657	AL107697 Drosophila
C	41	52.4	14.6	1101	17	CNS0177R	AL098462 Drosophila
C	42	52.4	14.6	1201	17	CNS01038	AL0880113 HS_4868_B
C	43	52.2	14.6	403	17	AG0880113	AL411257 TF end of
C	44	52.2	14.6	759	17	CNS060XV	BH148938 ENPPT52TF
C	45	52.2	14.6	765	17	BH148938	

ALIGNMENTS

RESULT 1	AL762543	176 bp	DNA	linear	GSS 19-JUN-2002
LOCUS	AL762543				
DEFINITION	Arabidopsis thaliana T-DNA flanking sequence GK-025603-013722,				
ACCESSION	AL762543				
VERSION	AL762543.1	GI:21508636			
KEYWORDS	GSS.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H. and Weisshaar,B.				
TITLE	A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines				
JOURNAL	Unpublished				
REFERENCE	Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.				
AUTHORS	A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 176)				
AUTHORS	Li,Y., Strizhov,N., Rosso,M. and Weisshaar,B.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer				

COMMENT

Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion within the locus defined by clone T3A5. The
 sequences are generated at the MPI for Plant Breeding Research in
 the context of the GABI-Kat project. GABI-Kat is part of the German
 plant Genomics program designated 'GABI'. Information on line
 availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

Location/Qualifiers

1..176

/organism="Arabidopsis thaliana"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-025G03-013722"

/note="PCR was performed on DNA from Arabidopsis thaliana
 plants (T1) which were transformed with the T-DNA from
 vector pAC106. The lines contain one or more T-DNA
 insertions. The DNA fragment(s) resulting from the PCR
 were directly sequenced to determine the genomic sequence
 flanking the insertion. Sequences displaying significant
 similarity to the A. thaliana nuclear genome sequence were
 processed for submission. T-DNA derived sequences were
 removed"

BASE COUNT 60 a 33 c 27 g 56 t
 ORIGIN

Query Match 38.4%; Score 137.4; DB 17; Length 176;
 Best Local Similarity 99.3%; Pred. No. 1e-11;
 Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 220 TTGATTATTAGTGCACACTTAACTTAAAGGCTCTCTAATAGCCCAATTATA 279

DB 1 TTGATTATTAGTGCACACTTAACTTAAAGGCTCTCTAATAGCCCAATTATA 60

OY 280 TTACGATTATACCAAGTGACAACTTTACTCGTTTGGATCGAAGCAATTAACAAT 339

DB 61 TTACGATTATACCAAGTGACAACTTTACTCGTTTGGATCGAAGCAATTAACAAT 120

OY 340 GTCAAAATACCAACACAG 358

DB 121 GTCAAAATACCAACACAG 139

RESULT 2

AL762544

LOCUS AL762544 194 bp DNA linear GSS 19-JUN-2002

DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-025G03-013757,
 genomic survey sequence.

ACCESSION AL762544

VERSION AL762544.1 GI:21508641

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
 This sequence is recovered from the left border of the T-DNA. It
 indicates an insertion within the locus defined by clone T3A5. The
 sequences are generated at the MPI for Plant Breeding Research in
 the context of the GABI-Kat project. GABI-Kat is part of the German
 plant Genomics program designated 'GABI'. Information on line
 availability can be found at:
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES

source

Location/Qualifiers

1..194

/organism="Arabidopsis thaliana"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-025G03-013757"

/note="PCR was performed on DNA from Arabidopsis thaliana
 plants (T1) which were transformed with the T-DNA from
 vector pAC106. The lines contain one or more T-DNA
 insertions. The DNA fragment(s) resulting from the PCR
 were directly sequenced to determine the genomic sequence
 flanking the insertion. Sequences displaying significant
 similarity to the A. thaliana nuclear genome sequence were
 processed for submission. T-DNA derived sequences were
 removed"

BASE COUNT 66 a 37 c 28 g 63 t
 ORIGIN

Query Match 35.9%; Score 128.4; DB 17; Length 194;
 Best Local Similarity 98.6%; Pred. No. 2.1e-10;
 Matches 140; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 217 TTGTTATTATTAGTGCACACTTAACTTAAAGGCTCTCTAATAGCCCAATT 276

DB 27 TTGTTATTATTAGTGCACACTTAACTTAAAGGCTCTCTAATAGCCCAATT 85

OY 277 ATATTGATTATTAACAAGTGACAACTTTACTCGTTTGGATCGAAGCAATTAACA 336

DB 86 ATATTGATTATTAACAAGTGACAACTTTACTCGTTTGGATCGAAGCAATTAACA 145

OY 337 ATTGTCAAAATACCAACACAG 358

DB 146 ATTGTCAAAATACCAACACAG 167

RESULT 3

AV787618/c

LOCUS AV787618 423 bp mRNA linear EST 28-MAR-2002

DEFINITION Arabidopsis thaliana cDNA clone RAF106-76-F19 3',
 mRNA sequence.

ACCESSION AV787618

VERSION AV787618.1 GI:19806408

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Large scale analysis of Arabidopsis full-length cDNA (2002b)
 Unpublished (2002)
 Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: mseki@fc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially
 as reported previously (Seki et al., 1998). This clone is in a

Location/Qualifiers

```
/clone="BACR48P1g"  
/clone_11b="RPCT-98"
```

BASE COUNT
ORIGIN

469 a

```

end : TET3
6 c 69 g

```

151 t	406 others
-------	------------

Query Match	19.0%	Score 68;	DB 17;	Length 1101;
Best Local Similarity	26.4%	Pred.	NO. 0.1;	
Matches	68.	Consensus	0.1;	

BASE COUNT	ORIGIN	156 a	79 c	64 g	124 t
156	a	79	c	64	g
124	t				

Query Match	33.8%;	Score 121;	DB 10;	Length 423;
Best Local Similarity	100.0%;			
Matches 121;	Consensus 100.0%;	Pred. No.	1.8e-09;	

1 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTCTTTTTCATTTAAATCA 60

61 TTTTTCGCCCCAATGATATATAAAAAATTGGATAAATAATATTATGGATAATCGTTTT 120

[illegible]

RESULT 4
NS00LT2

DEFINITION	1101 bp	DNA	linear	GSS 14-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR8P19 of RPCI-98 library from Drosophila melanogaster (fruit fly) (vector)				

VERSION	AL078714.1	GI:5102004
EXWORDS	GSS.	
SOURCE	Protein Data Bank	

Eukaryota: Metazoa: Arthropoda: Hexapoda: Insecta: Pterygota; Neoptera: Endopterygota; Diptera: Brachycera: Muscomorpha: Embryota: Eurytomidae

AUTHORS (please list all authors)
TITLE Genoscope, a genome browser
JOURNAL Direct Submission
CATEGORY Submissions

Correspondence: Dr. J. L. Viret, Institut Pasteur, 25 rue Docteur Roux, 75013 Paris, France. E-mail: secretf@genoscope.cns.fr
 Web: www.genoscope.cns.fr

melanogaster is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The RBCP *Drosophila*

Dr. Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by

recombinant strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, and the P1 and EST libraries is available at <http://www.flybase.org>.

source
location/Qualifiers
1. .1101

```
/db_xref="taxon:7227"  
/clone="BACR48P19"  
/clone_lib="RPCT-99"
```

BASE COUNT	a	c	g	t	others
ORIGIN	469	6	69	151	406

Best Local Similarity 26.48; Pred. NO. 0.1;
Matches 68; Conservative 100; Mismatches 90; Indels 0; Gaps 0;

[illegible]

```
Db . . . 663 TTTT TTTT TT AAAA TAA AAWT TD WAA AAWT TT KKKK KKAAD KKDAKKDGA 722
```

Db 723 KKAATTKKKKKAAWAAADKDDRRKGGGKKGKGGKKGGKKKKKKGGGKGGK 183

```

      :: :: :: :: :: | :: :: :: :: :: | :: :: :: :: :: | :: :: :: :: :: |
783 AGDDDAKDTKKKKKKAATTTT KKKKKKKKKKKKAADRTKTTKDAAAAAAAAK 842

```

```

      843  TKDKGKKKKKKTKKKKTTKK  860

```

CNS00DPKY
LOCUS

CONSULT 5
CNS00DPKY

ACCESSION
AL071865
BACR27A24 of RPCI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.

KEYWORDS
SOURCE
ORGANISM
GSS:
Drosophila melanogaster.
Drosophila melanogaster.

Neoptera; Endopterygota; hexapoda; Insecta; Pterygota; Ephyndroidea; Drosophilidae; Drosophila.

TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999)
Genoscope - Centre National de Séquençage
RP 191 9106 mmm

The RBCP is consistent with the Berkeley Drosophila Genome Project (rbcnp) determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (rbcnp) (http://www.soe.ucsf.edu/crms.ii) microscope.cms.ii)

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC Library was prepared by Kazutoyo Osogawa and Aaron Mammociani.

The library is named RPCR-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the

and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://www.bacpac.org>.

```
source  
    .928  
organism="Drosophila melanogaster"  
abundance=
```

BASE COUNT 262 a /note="end : 17"
ORIGIN 70 c 84 g 321 t 191 others

Query Match 18.3%; Score 65.4; DB 17; Length 928;
Best Local Similarity 31.0%; Pred. No. 0.27;
Matches 99; Conservative 83; Mismatches 137; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGAGTATTTATCTCTCTTTTGTGATATTTAATCA 60
DB 524 AAAAAAAAAATTTTWTWTATTTTAAATTTTWTWTATTTTAA 583
OY 61 TTTTTCGCCAATGATATATATTTGATATATATATATATATAT 120
DB 584 ATTTTATTTTAAATTTTATATATATATATATATATATATAT 643
OY 121 TTAGTCGGGTTTGAGAAAGGTTTCGACTTTCGAGATGATATAG 180
DB 644 WTATTTTATATATATATATATATATATATATATATATATAT 703
OY 181 GAGCTAGTTGAGTCTTGGACATTTGATGATGATGATGATGAC 240
DB 704 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 763
OY 241 ATTTAAGCTTAATGGCTTTCTATATAGCCCAATATATATAT 300
DB 764 WTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 823
OY 301 AACTTTTACTCTGTTTGTG 319
DB 824 KAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 842

RESULT 6

CNS00KHX 905 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACR17N06 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL077798.1 GI:4957174
VERSION AL077798
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 905)
AUTHORS Genoscope.

COMMENT Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)

COMMENT

JOURNAL

The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammeter in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. .905
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR17N06"

BASE COUNT 388 a /clone_lib="RPI-98"
ORIGIN 82 c 77 g 194 t 164 others

Query Match 16.8%; Score 60; DB 17; Length 905;
Best Local Similarity 35.9%; Pred. No. 1.8;
Matches 127; Conservative 59; Mismatches 168; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAGAGTATTTATCTCTCTTTTGTGATATTTAATCA 60
DB 882 AAAAAAAAAATTTTWTWTATTTTAAATTTTWTWTATTTTAA 823
OY 61 TTTTTCGCCAATGATATATATTTGATATATATATATATATAT 120
DB 822 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 763
OY 121 TTAGTCGGGTTTGAGAAAGGTTTCGACTTTCGAGATGATATAG 180
DB 762 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 703
OY 181 GAGCTAGTTGAGTCTTGGACATTTGATGATGATGATGATGAC 240
DB 702 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 643
OY 241 ATTTAAGCTTAATGGCTTTCTATATAGCCCAATATATATAT 300
DB 642 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 583
OY 301 AACTTTTACTCTGTTTGTG 354
DB 582 AHAAMMMMAATTTTWTWTATTTTAAATTTTAAATTTTAAAT 529

RESULT 7

CNS02N50/C 1027 bp DNA linear GSS 15-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
DEFINITION 163M16 of library G from Tetradon nigroviridis, genomic survey
sequence.

ACCESSION AL212733.1 GI:7871552
VERSION AL212733
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 1027)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fzames, C., Mincker, P., Brothier, P., Quetier, F.,
Saurin, W., and Weissenbach, J.

COMMENT Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence

REFERENCE

JOURNAL Unpublished
AUTHORS 2 (bases 1 to 1027)
Bouneau, L., Jallion, O., Dasilva, C., Fzames, C., Fisher, C.,
Bernot, A., Fzames, C., Mincker, P., Brothier, P., Quetier, F.,
Saurin, W., and Weissenbach, J.

TITLE

Characterization and repeat analysis of the compact genome of the
freshwater puterfish Tetradon nigroviridis
Unpublished
3 (bases 1 to 1027)
Genoscope.

COMMENT Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
genome. For more information project of the Tetradon nigroviridis
<http://www.genoscope.cns.fr/Tetradon>.

FEATURES
source
1. .1027
Location/Qualifiers

b
737 WAAAAAAAAATTKAYDTTTTTTTTWTWTWTWTBTITTAWTTTAAKDTTTTTTTW 796

ORIGIN	210 d	63 C	91 g	289 t
2222				
2221				
2220				
2219				
2218				
2217				
2216				
2215				
2214				
2213				
2212				
2211				
2210				
2209				
2208				
2207				
2206				
2205				
2204				
2203				
2202				
2201				
2200				
2199				
2198				
2197				
2196				
2195				
2194				
2193				
2192				
2191				
2190				
2189				
2188				
2187				
2186				
2185				
2184				
2183				
2182				
2181				
2180				
2179				
2178				
2177				
2176				
2175				
2174				
2173				
2172				
2171				
2170				
2169				
2168				
2167				
2166				
2165				
2164				
2163				
2162				
2161				
2160				
2159				
2158				
2157				
2156				
2155				
2154				
2153				
2152				
2151				
2150				
2149				
2148				
2147				
2146				
2145				
2144				
2143				
2142				
2141				
2140				
2139				
2138				
2137				
2136				
2135				
2134				
2133				
2132				
2131				
2130				
2129				
2128				
2127				
2126				
2125				
2124				
2123				
2122				
2121				
2120				
2119				
2118				
2117				
2116				
2115				
2114				

[illegible]

[illegible]

```

Db      685 AANTTTAAATTTTTTNNAAATNTTTTTTTTTTTTTTTGNNATNMAWAAAAAAMTWTTTTT 744
OY      70 CCCATGATATATAAAAATTGGATTAATATATATATGATTCGTCTTTTAGTCCG 129
       | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
Db      745 TTTTTTTTTTTTTTATATATATWMAWAAAAATWARAANAAGTATATTTTTTTTGGCGR 804
       : | | | | | : : : : : : : : : : : : : : : : : : : : : : : :
OY      130 GTTTGAGAAAAGGTTTGACCTTCGAAGAATGCAGATATATNAGATGGAGCTAGGT 189
       : | | | | | : : : : : : : : : : : : : : : : : : : : : : : :
Db      805 KATNAWAAAAAWMAACWAGATGDWWMAWAAAAATAGRTTWWWWTATATAATKRWHTDGGKG 864
OY      190 TGAGTCTTTGGACATTTGTATGTGATGTGGTGTGTTTATGTGCCA 236
       | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
Db      865 TTTTTTTTTTTTAKTTKGGGGGRITTTTKKKKKGGKKRAMWA 911

RESULT_13
LOCUS      CENS0029N                               1101 bp    DNA          linear   GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence Y7 end of BAC
            BACN01C10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AI097397
VERSION     AI097397.1 GI:5609008
KEYWORDS   GSS.
SOURCE      Drosophila melanogaster.
            Drosophila melanogaster
ORGANISM   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage ;
            BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr
            - web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billand at CEPR (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelobAC11.

FEATURES             Location/Qualifiers
     source           1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN01C10"
                     /clone_id="DrosBAC"
                     /plasmid="pBelobAC11"
                     /note="end : 77"
BASE COUNT        254 a      288 c      267 g      177 t      115 others
ORIGIN
Query Match      15.8%; Score 56.4; DB 17; Length 1101;
Best Local Similarity 47.4%; Pred. No. 5.7;
Matches      83; Conservative 24; Mismatches 68; Indels 0; Gaps 0;

OY      2 AAAAAAAAAAGATGAAGTATTTTATCTCTCTTTTGGATTAATTTAATCAT 61
       :| ||||| | : : : : | | | :| ||||| : | | : |
Db      892 WAAAAAAMAAAAAAMAAAAATTTTTTTTTTAAATTTTAAATTTAAAAA 951
       : | | | | | : : : : : : : : : : : : : : : : : : : : : : : :
OY      62 TTTTTTCCCACATATATAAATTTGGATAAATATATATTTGATATTCGTTTT 121
       ||||| : : : : : : : : : : | :||| | | : : : | ||||
Db      952 TTTTTCGATTTWWTTTNTKTADGATAMWAAATTTTWTATAATTTTTRKTT 1011
       ||||| : : : : : : : : : : | :||| | | : : : | ||||
OY      122 TAGTTCGGGTTGAAAAAGGTTTCGACTTCGGAAGTGACAGCTATATCA 176
       | | | | | : | | | : | | | : | | | : | : | : | : | : |
Db      1012 TTTTTTATTTWATGACARTTTTTTTTTTTWTATTAARAAARGAGMTTDAAMA 1066

```


